



APLICAÇÃO DE TÉCNICAS ESTATÍSTICAS PARA DIAGNÓSTICO DE ERROS EM MEDIDORES REDUNDANTES DE VAZÃO

Thayane Picorelli Correia

Projeto de Final de Curso

Orientadores

Príamo Albuquerque Melo Junior, D.Sc.

Thiago de Sá Feital, D.Sc.

Fevereiro de 2020

APLICAÇÃO DE TÉCNICAS ESTATÍSTICAS PARA DIAGNÓSTICO DE ERROS EM MEDIDORES REDUNDANTES DE VAZÃO

Thayane Picorelli Correia

Projeto de Final de Curso submetido ao Corpo Docente da Escola de Química, como parte dos requisitos necessários à obtenção do grau de Engenharia Química.

Aprovado por:

Carlos André de Vaz Júnior, D.Sc

Lízia Baleeiro Brack Corsetti

Orientado por:

Príamo Albuquerque Melo Junior, D.Sc

Thiago de Sá Feital, D.Sc

Rio de Janeiro, RJ - Brasil
Fevereiro de 2020

Correia, Thayane Picorelli.

Aplicação de técnicas estatísticas para diagnóstico de erros em medidores redundantes de vazão/ Thayane Picorelli Correia. Rio de Janeiro: UFRJ/EQ, 2020.

xiii, 55 p.; il.

(Monografia) - Universidade Federal do Rio de Janeiro, Escola de Química, 2020.

Orientadores: Príamo de Albuquerque Melo Junior e Thiago de Sá Feital.

1. Monitoramento Estatístico. 2. Análise em componentes principais. 3. Detecção de falhas. 4. Monografia. (Graduação - UFRJ/EQ). 5. Príamo de Albuquerque Melo Junior e Thiago de Sá Feital. I. Título.

AGRADECIMENTOS

Agradeço primeiramente a Deus e a seus intercessores pelas oportunidades que tive ao longo do caminho.

À minha família, em especial à minha irmã, mãe, avó Elza e ao meu pai, por terem me tornado mais forte e determinada.

Aos amigos Nayane, Ellen, Marcos e Louise, obrigada pela paciência e por compartilharem comigo os momentos bons e ruins.

Aos amigos da graduação, Julia, Thiago, Roberto e Matheus, obrigada por toda a ajuda nos períodos de desespero e alegria.

Aos meus orientadores, Príamo e Thiago, obrigada pela atenção, dedicação e por todos os ensinamentos.

À Lízia, por todo o aprendizado no período de estágio e pelo suporte na realização deste trabalho. Um agradecimento também ao profissional Felipe Khun pela disponibilidade e entusiasmo.

Agradeço a banca, Professores Carlos André e Ladimir, pela disponibilidade em participar da avaliação deste trabalho e pelas aulas durante a graduação.

Por fim, agradeço à Escola de Química da UFRJ, pelo ensino público, gratuito e de qualidade.

Resumo do Projeto Final de Curso apresentado à Escola de Química como parte dos requisitos necessários para obtenção de grau de Bacharel em Engenharia Química.

APLICAÇÃO DE TÉCNICAS ESTATÍSTICAS PARA DIAGNÓSTICO DE ERROS EM MEDIDORES REDUNDANTES DE VAZÃO

Thayane Picorelli Correia

Fevereiro, 2020

Orientadores: Prof. Príamo Albuquerque Melo Junior, D.Sc

Thiago de Sá Feital, D.Sc

A indústria de processos químicos necessita de um acompanhamento da confiabilidade dos medidores de vazão em função de questões relacionadas à segurança do processo, ao controle econômico e de *performance* e até mesmo de *compliance* industrial. Diversas técnicas estatísticas foram desenvolvidas ao longo dos anos para a realização deste acompanhamento. Este trabalho apresenta duas técnicas estatísticas, o Espectro de Variância e a Análise de Componentes Principais (PCA), visando obter uma análise da condição de três medidores redundantes de vazão da corrente de eteno produzida em uma planta petroquímica real. O principal objetivo é obter um procedimento para a detecção e o diagnóstico de falha online desses medidores; para isso além da aplicação das técnicas aos dados captados do processo real, foi realizada uma simulação com os dados na presença de um erro artificial conhecido. Com o espectro de variância não foi possível traçar uma estratégia de diagnóstico, mas este permitiu realizar uma estimativa do impacto econômico do erro de medição relativo a sinais de alta frequência. Já a PCA mostrou-se satisfatória para a detecção e diagnóstico de erros, mas com dois pontos de ressalva: o limite superior utilizado não se mostrou adequado para o erro simulado; e caso dois medidores apresentem erro, durante o diagnóstico pode ser interpretado que o único medidor correto é o maior contribuinte para o SPE (*Squared Prediction Error*).

ÍNDICE

CAPÍTULO I — INTRODUÇÃO	1
I.1 Motivação	1
I.2 Produção de eteno.....	1
I.3 Apresentação do processo	2
I.4 Identificação e diagnóstico de falhas.....	3
I.5 Detecção de erros grosseiros	5
I.6 Objetivo	6
CAPÍTULO II — REVISÃO BIBLIOGRÁFICA	7
CAPÍTULO III — METODOLOGIA.....	9
III.1 Espectro de Variância	9
III.2 PCA.....	13
CAPÍTULO IV - ANÁLISE DOS DADOS / DISCUSSÃO	19
IV.1 Visualização dos Dados	19
IV.2 Caracterização dos Dados	20
IV.2.1 Correlação entre os medidores A e B	20
IV.2.2 Correlação do medidor C com os outros medidores	21
IV.3 Espectro de Variância	23
IV.3.1 Análise com comprimento de janela grande	23
IV.3.2 Análise com comprimento de janela pequeno.....	25
IV.4 Análise de Componentes Principais	28
IV.4.1 Percentagem de variância cumulativa	28
IV.4.2 Variáveis originais e Variáveis latentes	29
IV.4.3 <i>Squared Prediction Error</i> (SPE).....	36

IV.4.4 Contribuições	38
IV.5 Simulação de erro de medição	42
CAPÍTULO V - CONCLUSÃO.....	50
V.1 Conclusões	50
V.2 Sugestões para Trabalhos Futuros	51
REFERÊNCIAS	53

ÍNDICE DE FIGURAS

Figura I-1 Corrente de eteno, com foco na transferência para a planta de polietileno.....	2
Figura I.2 Classificação dos métodos de detecção e diagnóstico de falhas.....	4
Figura III -1 Exemplificação dos tipos de janelas, , para um tamanho de janela igual a 5	10
Figura III-2 Variável medida ao longo do tempo, com as respectivas médias de cada janela disjunta de tamanho 10.....	11
Figura III-3 Variável em cada janela disjunta de tamanho 10, bem como a média destas variâncias, que representa a variância para o tamanho de janela 10.....	11
Figura III-4 Espectros de variâncias, gerado a partir de conjunto de dados para exemplificação. Apresentado espectro de variância construído a partir da mediana como medida de tendência central e o espectro gerado ao ser utilizada a média como medida de tendência central.....	12
Figura III-5 Exemplo de alguns possíveis novos eixos e as respectivas variabilidades capturadas.....	14
Figura III-6 Variáveis A e B plotadas a partir de uma base de dados de treinamento.....	16
Figura III-7 Novos eixos de coordenadas (variáveis latentes), obtidos através do critério de máxima variância, além dos eixos de coordenadas originais.....	16
Figura III- 8 Projeção de dado teste no subespaço principal, bem como posterior projeção no espaço original.....	17
Figura IV-1 Vazão de produção de eteno, em toneladas por hora, ao longo de cerca de um ano.	19
Figura IV-2 Vazão de produção de eteno, centralizada na média, ao longo de cerca de um ano.	20
Figura IV-3 Gráfico de correlação entre as vazões, em toneladas por hora, dos Medidores A e B.....	20
Figura IV-4 Gráfico de correlação entre as vazões [t/h] em patamares mais baixos dos Medidores A e B.....	21
Figura IV-5 Gráfico de correlação entre as vazões [t/h] em patamares mais elevados dos Medidores A e B.....	21

Figura IV-6 Gráfico de correlação entre a vazão, em toneladas por hora, do Medidor C com as vazões dos medidores A e B, na mesma unidade de medida	22
Figura IV-7 Gráfico de correlação entre a vazão em patamares mais baixos, em t/h, do Medidor C com as vazões dos medidores A e B, na mesma unidade de medida.....	22
Figura IV-8 Gráfico de correlação entre a vazão em patamares mais elevados, em t/h, do Medidor C com as vazões dos medidores A e B, na mesma unidade de medida.....	23
Figura IV-9 Espectro de Variância dos Medidores A, B e C, utilizando-se dados de vazão hora-a-hora ao longo de cerca de um ano.....	24
Figura IV-10 Percentuais, em relação ao desvio padrão máximo da amostra analisada, para os tamanhos de janelas percorridos em dias (eixo x).	24
Figura IV-11 Espectro de Variância dos Medidores A, B e C, utilizando-se dados de vazão hora-a-hora com o foco em janelas de pequenos tamanhos.	25
Figura IV-12 Espectro de Variância dos Medidores A, B e C, utilizando-se dados de vazão minuto-a-minuto ao longo de quatro dias.....	26
Figura IV-13 Espectro de Variância dos Medidores A, B e C, utilizando-se dados de vazão minuto-a-minuto com o foco em janelas de pequenos tamanhos.	26
Figura IV-14 Percentagem de variância cumulativa das três variáveis latentes reveladas pela PCA.	29
Figura IV-15 Percentagem de variância cumulativa das variáveis latentes, não caracterizadas como variáveis principais, conforme reveladas pela PCA	29
Figura IV-16 Visão inicial ao longo do tempo (dias) das três variáveis observadas (lado esquerdo) – em t/h – e das três variáveis latentes (lado direito).....	30
Figura IV-17 Visão ao longo do tempo (dias) das três variáveis observadas (lado esquerdo) – em t/h – e das três variáveis latentes (lado direito).....	31
Figura IV-18 Visão ao longo do tempo das três variáveis observadas (lado esquerdo) – em t/h – e das três variáveis latentes (lado direito) no período de 50 dias corridos.....	32
Figura IV-19 Visão ao longo do tempo das três variáveis observadas (lado esquerdo) – em t/h – e das três variáveis latentes (lado direito) no período de 7 dias corridos.....	32
Figura IV-20 Visão ao longo do tempo (dias) das três variáveis observadas (lado esquerdo) – em t/h – e das três variáveis latentes (lado direito), com foco no período de transição entre os dados de treinamento e os dados de teste.....	33

Figura IV-21 Visão ao longo do tempo (dias) das três variáveis observadas (lado esquerdo) – em t/h – e das três variáveis latentes (lado direito), com foco no período de transição entre os dados de treinamento e os dados de teste.....	34
Figura IV-22 Perfil de temperatura (°C) ao longo do ano na região metropolitana do Rio de Janeiro. Construído através dos dados de Climatempo (2020).	34
Figura IV-23 Perfil de temperatura média (°C) ao longo de 01 de janeiro.	35
Figura IV-24 Perfil de temperatura média (°C) ao longo de 01 de julho.	36
Figura IV-25 SPE e limite de erro calculado conforme literatura.....	37
Figura IV-26 SPE e limite de erro calculado conforme literatura.....	37
Figura IV-27 SPE e limite de erro calculado conforme literatura, com foco no período de transição entre os dados de treinamento e os dados de teste	38
Figura IV-28 Visão inicial ao longo do tempo (dias) das contribuições dos três medidores no erro, utilizando-se dados para treinamento e dados para teste (a partir do dia 300).	39
Figura IV-29 Visão ao longo do tempo (dias) das contribuições dos três medidores no erro, utilizando-se dados para treinamento e dados para teste (a partir do dia 300).....	39
Figura IV-30 Visão ao longo do tempo (dias) do percentual de cada contribuição para o SPE, utilizando-se dados para treinamento e dados para teste (a partir do dia 300).....	40
Figura IV-31 Visão ao longo do tempo (dias) do percentual de cada contribuição para o SPE, no período de 50 dias corridos.....	40
Figura IV-32 Visão do percentual de cada contribuição para o SPE, entre os dias 35 e 55..	41
Figura IV-33 Visão das contribuições dos três medidores no erro, com foco no período de transição entre os dados de treinamento e os dados de teste	41
Figura IV-34 Visão do percentual de cada contribuição para o SPE, com foco no período de transição entre os dados de treinamento e os dados de teste	42
Figura IV-35 Visão inicial ao longo do tempo (dias) das três variáveis observadas (lado esquerdo) – em t/h – e das três variáveis latentes (lado direito). O erro foi imputado no Medidor A a partir do dia 317	43
Figura IV-36 Visão ao longo do tempo (dias) das três variáveis observadas/estimadas (lado esquerdo) – em t/h – e das três variáveis latentes (lado direito). O erro foi imputado nos dados para teste do Medidor A a partir do dia 31	43

Figura IV-37 Visão ao longo do tempo (dias) das três variáveis observadas/estimadas (lado esquerdo) – em t/h – e das três variáveis latentes (lado direito), com foco no período de transição entre os dados de treinamento e os dados de teste (a partir do dia 300), imputando um erro no Medidor A a partir do dia 317.....	44
Figura IV-38 Visão ao longo do tempo (dias) das três variáveis observadas/estimadas (lado esquerdo) – em t/h – e das três variáveis latentes (lado direito), com foco no período de transição entre os dados de treinamento e os dados de teste (a partir do dia 300), imputando um erro no Medidor A a partir do dia 317.....	44
Figura IV-39 Visão inicial ao longo do tempo (dias) das contribuições dos três medidores no erro, utilizando-se dados para treinamento e dados para teste (a partir do dia 300), com erro conhecido no Medidor A a partir do dia 317.....	45
Figura IV-40 Visão ao longo do tempo (dias) das contribuições dos três medidores no erro, utilizando-se dados para treinamento e dados para teste (a partir do dia 300), com erro conhecido no Medidor A a partir do dia 317.....	45
Figura IV-41 Visão das contribuições dos três medidores no erro, com foco na transição entre os dados para treinamento e os dados para teste (a partir do dia 300), com erro conhecido no Medidor A a partir do dia 317	46
Figura IV-42 Visão inicial ao longo do tempo (dias) das contribuições dos três medidores no erro, utilizando-se dados para treinamento e dados para teste (a partir do dia 300), com erro conhecido no Medidor A a partir do dia 317.....	46
Figura IV-43 Visão ao longo do tempo (dias) das contribuições dos três medidores no erro, utilizando-se dados para treinamento e dados para teste (a partir do dia 300), com erro conhecido no Medidor A a partir do dia 317.....	47
Figura IV-44 Visão das contribuições dos três medidores no erro, com foco na transição entre os dados para treinamento e dados para teste (a partir do dia 300), com erro conhecido no Medidor A a partir do dia 317	48
Figura IV-45 Visão ao longo do tempo (dias) do percentual de cada contribuição para o SPE, utilizando-se dados para treinamento e dados para teste (a partir do dia 300), com erro no Medidor A imputado a partir do dia 317	48

Figura IV-46 Visão do percentual de cada contribuição para o SPE, com foco na transição entre os dados para treinamento e dados para teste (a partir do dia 300), com erro conhecido no Medidor A a partir do dia 317 49

ÍNDICE DE TABELAS

Tabela IV-1 Desvios padrão, em toneladas por hora, observados com tamanho de janela igual a 2 para os Medidores A, B e C, utilizando bases de dados estruturadas de formas diferentes	26
Tabela IV-2 Temperaturas médias típicas mensais na região metropolitana do Rio de Janeiro.	33

NOMENCLATURA

LISTA DE ABREVIATURAS

Sigla	Descrição
KPI	<i>Key Performance Indicator</i>
PCA	Análise de Componentes Principais
SPE	<i>Squared Prediction Error</i>

CAPÍTULO I — INTRODUÇÃO

I.1 Motivação

A indústria de processos químicos necessita de um acompanhamento da confiabilidade dos medidores de vazão em função de questões relacionadas à segurança do processo, ao controle econômico e de *performance* e até mesmo de *compliance* industrial.

Alguns instrumentos, especialmente os de vazão e nível, possuem um impacto importante em KPI's (*Key Performance Indicators*), sejam eles de eficiência ou eco-indicadores; um erro nesse aspecto, pode afetar os resultados ou mascarar a ocorrência de possíveis problemas. Além disso, erros de medição podem impactar no resultado econômico da empresa, tendo em vista que muitas das medições são utilizadas para contabilizar transferências entre áreas e entre empresas (compra e venda de correntes), se caracterizando como medições fiscais.

No tocante à segurança, uma indicação de medição falsa não identificada pode acarretar ações desnecessárias ou inapropriadas, que levem a uma redução de eficiência ou até a um problema operacional no processo.

Por isso, é importante ter uma ferramenta de acompanhamento do desempenho desses instrumentos e de detecção de eventuais erros na medição, de preferência que consiga capturar não somente erros grosseiros de alta magnitude, mas também erros mais suaves. O desenvolvimento de tal ferramenta possibilitaria detecções mais ágeis de erros de medição, tornando a ação de verificação e correção em campo mais imediata, e com isso potencialmente diminuindo o período de erro, minimizando os impactos.

No contexto industrial de uma planta de produção de eteno foi identificada uma oportunidade de identificação de erros em instrumentos de vazão, a partir da comparação dos perfis gerados pelas quantidades medidas ao longo do tempo, tendo disponíveis dois ou mais medidores redundantes.

I.2 Produção de eteno

Dentre as principais resinas consumidas no Brasil pela indústria de transformação, o polietileno possui a maior parcela de consumo, totalizando cerca de 31,9%, percentual

distribuído entre polietileno de alta densidade (13,6%), polietileno de baixa densidade linear (10,4%) e polietileno de baixa densidade (7,9%). [Informação obtida em: <http://file.abiplast.org.br>, 21 jan. 2020]

A matéria-prima para produção de polietileno é o eteno. Em 2017, a capacidade global de produção de eteno era de cerca de 170 milhões de toneladas/ano e dentre essa quantidade, cerca de 50% foi produzida pelas 10 maiores indústrias do segmento. Em 2017, a Braskem se localizava em nono lugar do ranking, sendo responsável pela produção de cerca de 3% do total, o que correspondia a 5 milhões de toneladas anuais. [Informação obtida em: <https://www.braskem.com.br/RI/o-setor-petroquimico>, acesso, 21 jan. 2020]

I.3 Apresentação do processo

Neste trabalho, foi considerada uma planta de produção de eteno, que utiliza etano e propano como matéria-prima e além do eteno, que é o produto principal, produz subprodutos, como propeno e mais pesados. O foco do trabalho foi nos medidores de vazão de eteno para a planta de segunda geração, tendo em vista que este é o produto principal da unidade de negócio, que impacta tanto no fechamento do balanço de massa quanto também no indicador de eficiência da planta, além dele ser uma corrente de transferência.

Abaixo, segue descrição resumida do processo de acordo com o que pode ser encontrado nos trabalhos de Souza e Mustafa (2017) e Hombееck (2019) e ilustração com foco na parte estudada. Conforme pode ser observado pela Figura I-1, os Medidores A e C medem a mesma corrente, assim como o somatório dos Medidores B1 e B2, nomeado neste trabalho como Medidor B.

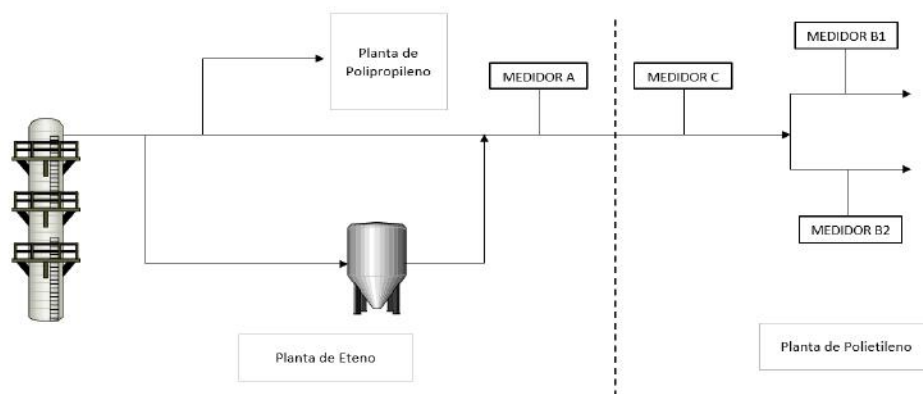


Figura I-1 Corrente de eteno, com foco na transferência para a planta de polietileno.

O etano e propano gás são craqueados nos fornos de pirólise a corrente resultante passa por sistemas de compressão, secagem e resfriamento. No trem de resfriamento, há a recuperação de hidrogênio. O gás comprimido segue então para a desmetanizadora, separando o metano e mais leves do resto da corrente.

A corrente de C₂⁺ segue para uma coluna de desetanização (destiladora que separa eteno, etano e etino no topo e C₃⁺ no fundo). A corrente de C₂ (topo) é separada na destiladora de eteno, retirando eteno no topo e etano no fundo, que é recirculado no processo. O eteno que sai no topo passa por reatores de acetileno, com o objetivo de converter eventuais moléculas de etino em eteno. Após passar pelos reatores, essa corrente de eteno pode ter alguns destinos, são eles: armazenamento em tanque pulmão, envio para a planta de produção de polietileno, envio para a planta de polipropileno ou a combinação destes ao mesmo tempo. Cabe ressaltar que a o eteno armazenado em tanque também pode ser transferido para as linhas de consumo das plantas da segunda geração. Os demais produtos têm destinação para clientes externos ou para outras unidades industriais, de acordo com a melhor estratégia econômica para a empresa.

A corrente de fundo da desetanizadora, rica em C₃⁺, segue para a despropanizadora (coluna de destilação que separa o propano, propeno, propino e propadieno no topo e C₄⁺ no fundo). A corrente de C₃ (topo) é separada na destiladora de propeno, retirando propano no fundo e propeno no topo, que antes de ser direcionado para a planta de polipropileno passa por um reator de MAPD (metilacetileno propadieno) com o objetivo de maximizar a conversão de propino e propadieno em propeno.

A corrente de C₄⁺ (fundo da despropanizadora) é separada na desbutanizadora (coluna de destilação), saindo no topo a corrente de butano e no fundo C₅⁺, que segue para a área de fracionamento de gasolina.

I.4 Identificação e diagnóstico de falhas

O campo de detecção e diagnóstico de falhas tem uma importância expressiva para o controle de processos industriais, por isso tem sido amplamente explorado. Abordagens alternativas aos métodos clássicos de detecção e diagnóstico de falhas foram propostas. Estas alternativas são baseadas em estatística multivariada, possibilitando projetar a maioria das informações importantes em espaço de menor dimensionalidade, onde regiões de operação

normal e anormal são mais facilmente identificadas. Conforme novos dados surgem, eles são comparados com os dados de variação normal do processo através dos modelos de dimensão reduzida e são utilizados testes estatísticos para detectar comportamentos anormais. Os modelos padrão são construídos facilmente através dessas abordagens, sendo somente necessária uma base de dados normais do processo (Yoon e MacGregor, 2012).

Para a seleção de um método de detecção e diagnóstico de falhas, algumas características desejadas devem ser levadas em consideração: detecção e diagnóstico rápidos, isolabilidade, robustez e adaptabilidade, tendo em vista que processos industriais reais podem ser mudados ao longo do tempo de operação e que se deve minimizar os falsos-positivos (Venkatasubramanian *et al.*, 2003).

Venkatasubramanian *et al.* (2003) classifica os métodos de detecção de falhas em três categorias principais, conforme Figura I.2: métodos quantitativos baseados em modelos, métodos qualitativos baseados em modelos e métodos baseados no histórico do processo. E somente para este último não é obrigatoriamente necessário ter um conhecimento fenomenológico do processo.

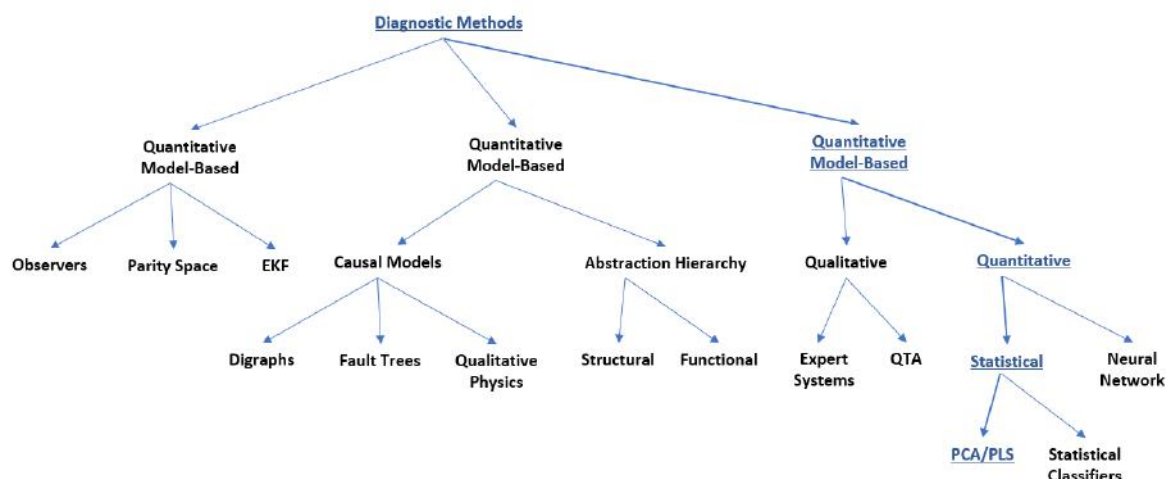


Figura I.2 Classificação dos métodos de detecção e diagnóstico de falhas. Baseado em Venkatasubramanian, 2003.

Os métodos baseados em histórico do processo, conhecidos como *data-driven*, têm sido massivamente estudados, tendo em vista a evolução computacional, que possibilitou a disponibilidade de grandes quantidades de dados do processo (Yin *et al.*, 2012).

Para a etapa de isolamento da falha, normalmente são utilizadas as parcelas de contribuição que representam a contribuição de cada variável no desvio identificado em relação ao modelo normal de operação. Essas parcelas são bastante eficazes em isolar falhas em sensores (Yoon e MacGregor, 2012).

Ainda de acordo com Venkatasubramanian *et al.* (2003), uma das formas de erro no processo acontece em atuadores e sensores, podendo ser erros fixos, relacionado a um viés constante ou erros não-sistêmicos. Dentro do contexto de controle de processos, o principal objetivo de se fazer o diagnóstico de erros em sensores é a rápida atuação sobre um instrumento para que este não prejudique seriamente o desempenho do sistema de controle. Com isso, intimamente ligada à área de detecção e diagnóstico de falhas, há a área de detecção de erros grosseiros nos dados dos sensores e validação dos instrumentos. Por isso, a validação e retificação de dados é um caso específico de um problema mais geral de diagnóstico de falhas.

I.5 Detecção de erros grosseiros

As medidas recebidas de um processo são inevitavelmente afetadas por erros durante a medição, processamento e transmissão. O erro total de uma medida pode ser compreendido como a soma de 3 tipos de erros (Marotta *et al.*, 2007):

- erros sistemáticos: aqueles geralmente de causa conhecida, podendo ser corrigidos;
- erros acidentais: geralmente de causa desconhecida até então e apresentando-se diferente em cada observação – influência aleatória nas medições – sendo necessário tratá-los com base na teoria da probabilidade;
- erros grosseiros: associados a eventos não aleatórios, por exemplo: necessidade de calibração do instrumento, mal funcionamento ou falha nos instrumentos, mudanças nas condições ambientais e/ou vazamento desconhecido no processo. Muitas vezes somente podem ser detectados através de testes estatísticos, tendo em vista que, em certos casos, quando os erros não são tão expressivos, sua detecção fica dificultada.

Em função da manutenção do desempenho dos sistemas de controle e da operação do processo, é importante minimizar os erros aleatórios e grosseiros captados pelos instrumentos (Barbosa, 2008).

A redundância é a única forma de contrastar dados e determinar potenciais falhas, por meio da filtragem destes dados para detectar vieses. À essa abordagem foi dado o nome de detecção de erros grosseiros (*gross error detection*) e mais recentemente de validação de dados ou de sensores (*data or sensor validation*) e reconstrução de sinal (*signal reconstruction*) (Barbosa, 2008).

A identificação de erros grosseiros pode ser feita com a aplicação de diversas técnicas, normalmente baseadas no teste estatístico de hipóteses, tais como Teste Global e Teste da Medição. A técnica PCA (Análise de Componentes Principais) foi aplicada à detecção de erros grosseiros pela primeira vez em 1995. Desde então outras técnicas também foram utilizadas para esse fim (Barbosa, 2008).

I.6 Objetivo

Este trabalho tem por objetivo principal aplicar técnicas estatísticas para resolver o problema da ausência de uma metodologia de diagnóstico *online* de erros de medição, a fim de, comprovada a adequação das metodologias propostas, aplicá-las na prática industrial. As técnicas estatísticas dispostas no presente trabalho como potenciais para entendimento e resolução desse problema foram: espectro de variância e análise de componentes principais (PCA).

Dentre os objetivos específicos estão:

- Caracterização inicial dos medidores de um sistema e comprovação do comportamento correlato entre eles;
- Análise dos medidores através dos espectros de variância dos mesmos, a fim de comparar o impacto de ruídos e erros de alta frequência nas respectivas medições;
- Exploração dos medidores e dos sinais do processo e de interferências através da técnica PCA;
- Validação da técnica de PCA para a resolução do problema descrito, ao simular um erro mapeado no conjunto de dados e analisar o *output* do método.

CAPÍTULO II — REVISÃO BIBLIOGRÁFICA

Os métodos básicos de detecção de falhas baseados em dados são análise de componentes principais (PCA – *Principal Componente Analysis*), mínimos quadrados parciais (PSL – *Partial Least Squares*), análise de componentes independentes (ICA – *Independent Component Analysis*), abordagem auxiliada do subespaço (SAP – *Subspace Aided Approach*) e análise do discriminante de Fisher (FDA – *Fisher Discriminant Analysis*) (Yin *et al.*, 2012).

Em relação à técnica de PCA, há várias aplicações relatadas na literatura atualmente, assim como combinações desta com outras técnicas e procedimentos, visando o aumento da eficácia na detecção e diagnóstico de falhas nos mais diversos cenários. Dentre os trabalhos disponíveis, alguns podem ser destacados, tais como: a adaptação do PCA para monitoramento de processo multimodal, capturando a relação não gaussiana e não linear do conjunto de dados, realizada por Zhang *et al.* (2012); diagnóstico de falhas de processo utilizando PCA e treinamento de modelo com análise externa, tendo em vista operações que ocorrem em múltiplas regiões, como descrita em Assis *et al.* (2017); utilização de WARP (*Weighted Adaptive Recursive PCA*) para minimizar alarmes falsos no monitoramento em função da dinâmica natural do processo, trabalhado por Portnoy *et al.* (2016); e adaptação da PCA para monitoramento de toda a planta, realizando monitoramento em bloco, sem a necessidade de conhecimento do processo para a divisão dos blocos (Ge e Song, 2013).

Normalmente, os estudos com PCA são mais amplamente relacionados à detecção de falhas no processo, mas já há diversos artigos que estudam a utilização desta técnica para detecção de falhas em sensores, como pode ser relatado em Harkat *et al.* (2006) (detecção e diagnóstico de sensores de monitoramento de ar utilizando a PCA), além de outras aplicações também relacionadas a sensores, como por exemplo em Qin *et al.* (1997) (abordagem de sensor de auto-validação com base na análise de componentes principais).

No âmbito mais amplo da área de detecção e diagnóstico de falhas, além da PCA, diversos estudos têm sido elaborados utilizando redes neurais, tais como em Lerner (2018), que utilizou redes neurais artificiais para detecção de falhas no processo *Tennessee Eastman* e em Loca e Rauber (2019), que explorou o uso de uma rede neural convolucional unidimensional para detecção de falhas em processos industriais. Além disso, é possível

encontrar na literatura a aplicação de diversas outras técnicas para detecção de falhas, como pode ser visto nos trabalhos enunciados a seguir: detecção de falhas no processo de benchmark do *Tennessee Eastman* com análise de espectro singular não linear (Krishnannair e Aldrich, 2017); modelos autoregressivos modificados descentralizados para detecção de falhas em processos dinâmicos (Tong *et al.*, 2018); autoencodificadores esparsos empilhados dinâmicos e com informações mútuas para detecção de falhas (Yin e Yan, 2019); detecção de falhas aprimorada com base em projeções de preservação global-local com análise quantitativa de estrutura global-local (Zhan *et al.*, 2017); e detecção e identificação de falhas recursivas para processos de variação temporal, no qual há análise de variação canônica (Shang *et al.*, 2016), dentre tantos outros trabalhos que podem ser encontrados na literatura que aborda o tema de detecção e diagnóstico de falhas.

As técnicas de reconciliação de dados e detecção de erros grosseiros de medidas começaram a ser exploradas na engenharia química na década de 1960. As metodologias para detecção de erros grosseiros evoluíram com o tempo e atualmente, além da técnica de PCA, técnicas baseadas em redes neurais têm sido estudadas (Barbosa, 2008).

Em relação ao espectro de variância, existem poucos relatos na literatura de sua aplicação para detecção de falhas em processos químicos, mas pode-se destacar os trabalhos de Feital e Pinto (2015) e Feital *et al.* (2012), visando a validação das medições de processos contínuos e objetivando estimar as variâncias dos erros de medição em processos industriais, respectivamente.

CAPÍTULO III — METODOLOGIA

O presente trabalho foi baseado em dados reais de uma petroquímica, captados ao longo de 1 ano de processamento. As informações originais foram obtidas em períodos de 1 hora, cujo valor apresentado é representado pela média da vazão nesse período de tempo e também foram coletados dados minuto-a-minuto para outras análises. A metodologia adotada é baseada na descrição de implementação de espectros de variância e PCA propostas, respectivamente, por Feital *et al.* (2015) e por Feital (2011), Bastidas (2018) e Qin (2003). A programação computacional das técnicas foi realizada em linguagem Python.

III.1 Espectro de Variância

A partir de uma base de dados, composta por m medidas de n variáveis, é realizado o agrupamento dos dados em quantidades pré-determinadas (tamanho de janela), e dentro dessas janelas é utilizada uma medida de tendência central para cálculo da variância. No caso deste trabalho, a base de dados é composta por medições de vazão ao longo do tempo e foram realizados testes utilizando a média e a mediana como medidas de tendência central dos dados. A partir das variâncias de cada uma das janelas de mesmo tamanho, que percorrem todo o conjunto dos dados, é calculada a variância média para um determinado tamanho de janela. Esse procedimento é realizado do tamanho de janela 2 até um tamanho de janela que englobe toda a quantidade de dados estudada.

Além disso, existem diversos tipos de janelas, que são elas: adjuntas, disjuntas e variações entre elas. As janelas disjuntas não possuem elemento em comum, já nas janelas adjuntas, os agrupamentos adjacentes possuem todos os dados, com exceção de um dado, em comum. No presente trabalho, foram utilizadas as janelas adjuntas, pois estas geram espectros mais suaves que as disjuntas e havia capacidade computacional para trabalhar com este tipo de janela. A Figura III-1 ilustra os dois tipos extremos de janelas.

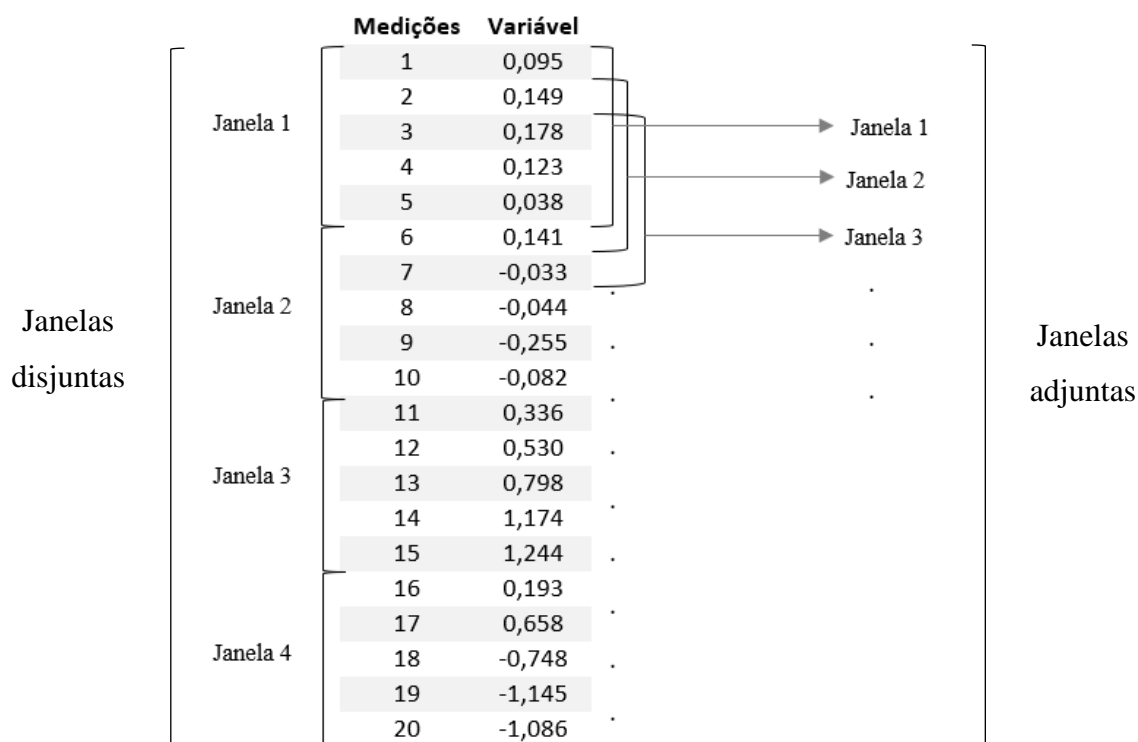


Figura III-1 Exemplificação dos tipos de janelas, para um tamanho de janela igual a 5.

Os gráficos abaixo foram desenhados para ilustrar o procedimento realizado na técnica de espectro de variância. As imagens III-2 e III-3 exemplificam o procedimento para janelas disjuntas de tamanho 10. A imagem III-4 corresponde ao espectro de variância gerado com a mesma base de dados, mas para janelas adjuntas, utilizando-se a média e a mediana como medida de tendência central.

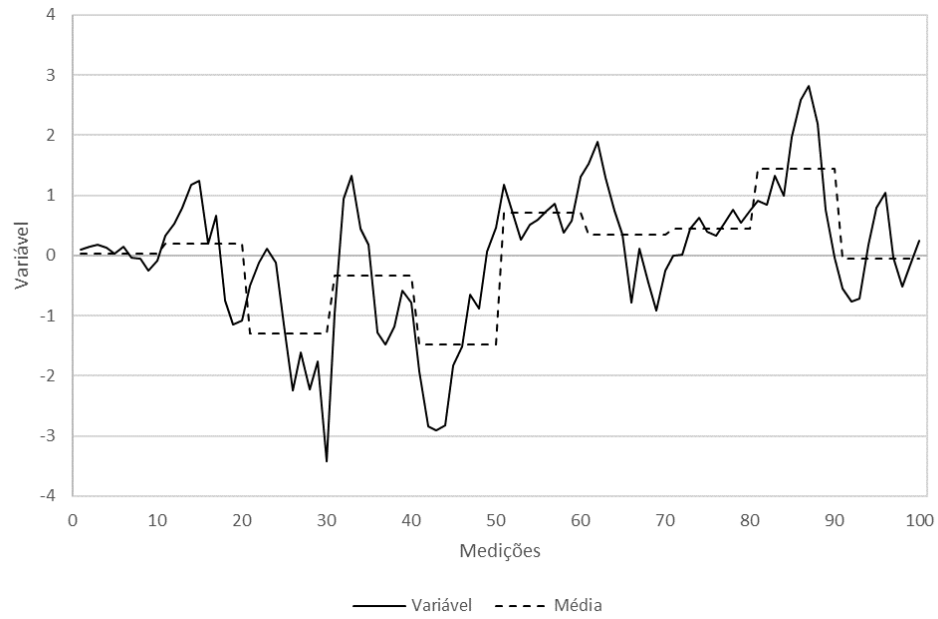


Figura III-2 Variável medida ao longo do tempo, com as respectivas médias de cada janela disjunta de tamanho 10.

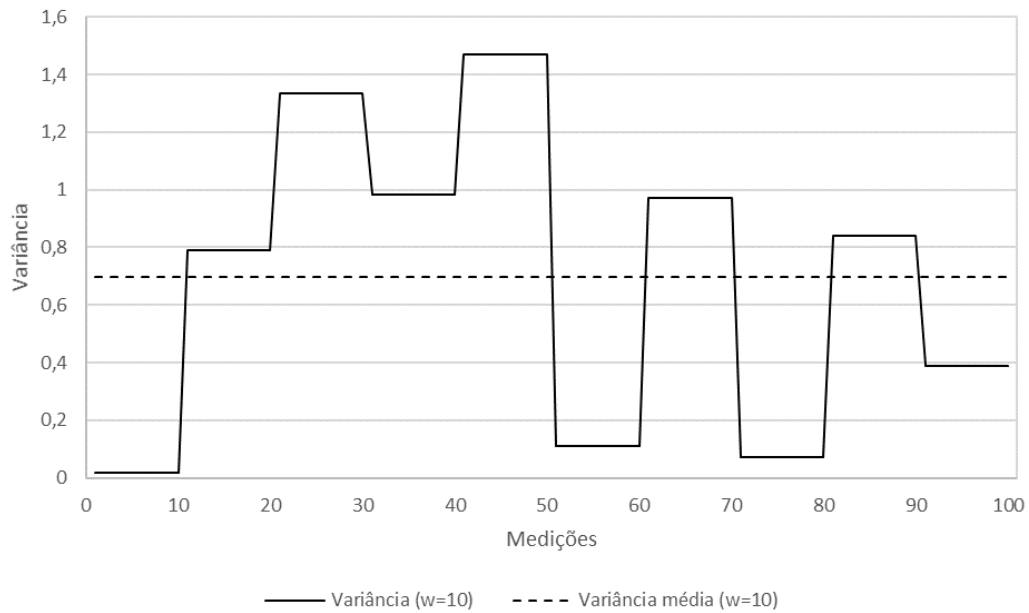


Figura III-3 Variável em cada janela disjunta de tamanho 10, bem como a média destas variâncias, que representa a variância para o tamanho de janela 10.

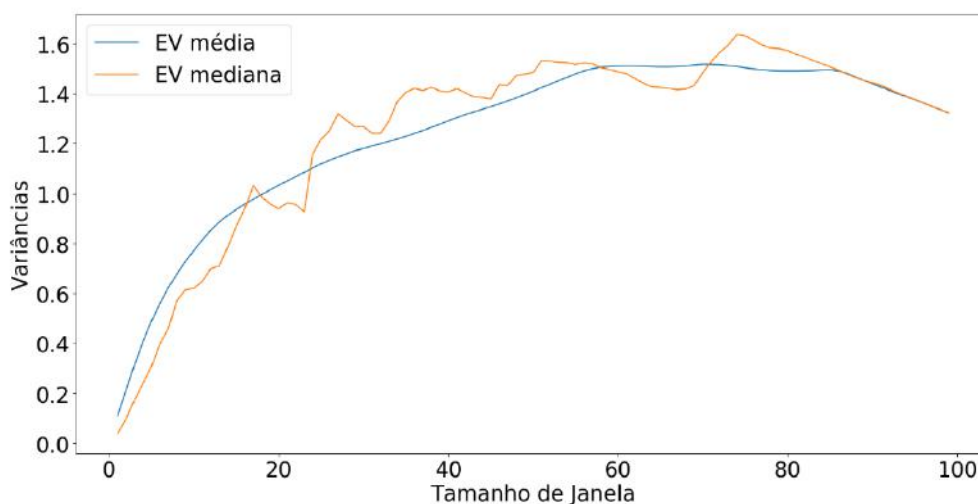


Figura III-4 Espectros de variâncias, gerado a partir de conjunto de dados para exemplificação. Apresentado espectro de variância construído a partir da mediana como medida de tendência central e o espectro gerado ao ser utilizada a média como medida de tendência central.

São utilizadas medições ordenadas no tempo para que os dados próximos entre si possuam características em comum, já que dados não ordenados poderiam gerar variações abruptas ou mais suavizadas, não decorrentes de um processo normal de elevação ou queda de vazão, o que seria traduzido em uma variância não representativa do medidor.

Para um tamanho de janela 2, espera-se que a variância seja a representação de somente erros de alta frequência que influenciem o instrumento, por exemplo, impactos na medição relativos à ruídos e tremulações e à própria precisão intrínseca do medidor. Ou seja, a variância para o tamanho de janela 2 poderia ser interpretada como uma medida indicativa do desempenho do sistema de medição analisado. Mas para que tudo isso seja verdadeiro, o tempo de amostragem tem que muito menor que a constante de tempo característica do processo, pois neste caso, as duas medidas consecutivas podem ser consideradas como réplicas da mesma medida.

Conforme o tamanho de janela aumenta, a dinâmica do processo é capturada, de forma que quando o tamanho de janela é igual ao número de observações da variável, é obtida a variância total do conjunto de dados em análise. Sendo assim, conforme a variação no comprimento de janela ocorra, potencialmente são revelados os ciclos de operação, a variabilidade diária e mudanças no desempenho dos instrumentos.

Como a variância não possui unidade física, no presente trabalho, de forma a obter interpretações físicas dos valores, bem como avaliar seu impacto econômico, foram gerados espectros com o desvio padrão (raiz quadrada da variância), que também é uma estatística que representa a variabilidade dos dados.

III.2 PCA

A análise de componentes principais (PCA) é uma técnica de redução de dimensionalidade, amplamente aplicada para monitoramento de processos, simplificando-o. Através dessa técnica, é obtido um novo conjunto de variáveis independentes/não correlacionadas, chamadas variáveis latentes (Bastidas, 2018).

As variáveis observadas possuem informações intrínsecas ao processo, informações das vizinhanças e do próprio sistema de medição. O PCA é utilizado para que essas contribuições sejam maximizadas individualmente através das diferentes variáveis latentes. Fazendo-se isso, é possível separar a contribuição do processo das contribuições da vizinhança e dos erros de medição, podendo-se obter assim um valor sem a influência dos erros e de efeitos externos ao processo. E com a seleção das direções principais é gerado um modelo do processo (Feital, 2011).

É importante ressaltar que, para ser utilizado o PCA, as variáveis estudadas devem possuir correlação entre si, ou seja, as contribuições mais importantes de processo, traduzidas nas suas medidas observadas, devem ser as mesmas para todas elas, para que a PCA possa realçar esses sinais intrínsecos (Feital, 2011).

Antes de enumerar as etapas da técnica de PCA, deve-se destacar que o critério de máxima variância é utilizado para a determinação da mudança de base. Ou seja, o novo eixo selecionado pela técnica deverá ser aquele onde há a maior variância possível em relação aos dados originais, isto é, onde é possível capturar a maior variabilidade dos dados (Feital, 2011). A Figura III-5 exemplifica esse critério; nela, demonstra-se dois possíveis eixos, dentre tantos outros, destacando-se a variabilidade em relação aos dados originais capturada por cada um deles. Neste caso, o eixo destacado em azul seria o escolhido pela técnica dentre os demais potenciais eixos, tendo em vista sua maior variabilidade. O segundo eixo será obtido por diferença, já que os eixos precisam ser ortogonais.

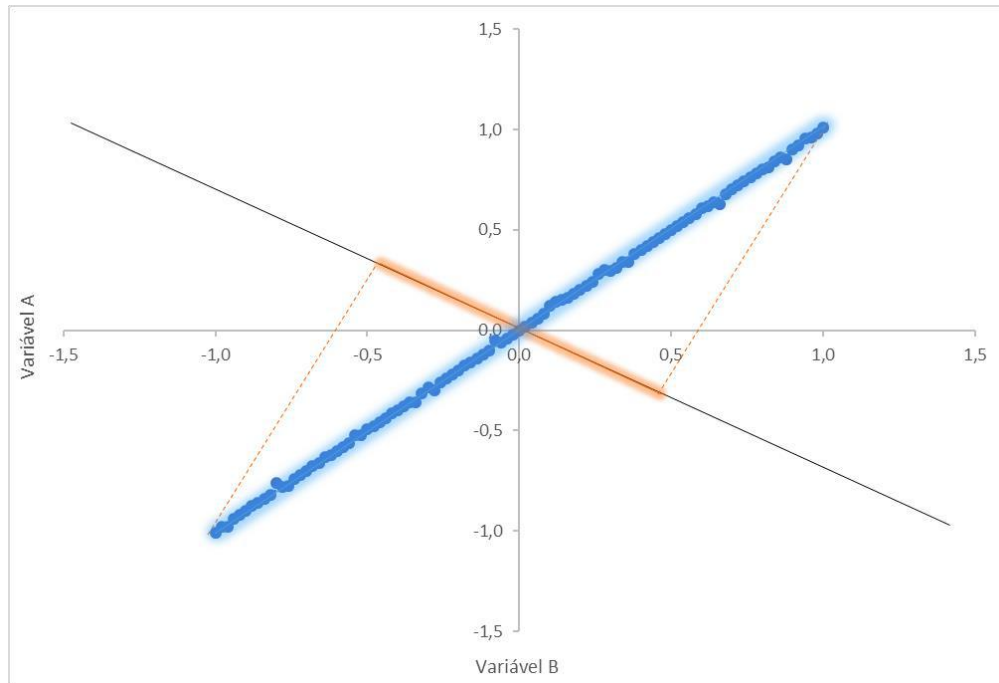


Figura III-5 Exemplo de alguns possíveis novos eixos e as respectivas variabilidades capturadas.

A PCA desempenha sua funcionalidade ao realizar a mudança de base das medidas observadas, ou seja, há uma mudança dos eixos de coordenadas (as coordenadas originais seriam as variáveis observadas). Essa mudança ocorre através da realização das seguintes etapas, que consiste na técnica de PCA (Bastidas, 2018):

- 1) Criação da base de dados, com n observações de cada uma das m variáveis (matriz X)

$$X = [x_1 \ x_2 \ \dots \ x_n]$$

- 2) Normalização dos dados contidos na base montada, ou seja, os dados devem ser centralizados nas médias. (Para que seja possível trabalhar com variáveis com magnitudes diferentes ao mesmo tempo)
- 3) A partir da matriz X com dados normalizados é calculada a matriz de covariância S .

$$S_{XX} = \frac{X^T X}{n - 1}$$

- 4) Cálculo de autovetores e autovalores do sistema a partir da matriz S . Com os autovalores, são calculadas as variâncias de cada um dos componentes, para se

fazer a Percentagem de Variância Cumulativa, de forma a identificar as componentes principais.

$$S_{XX} v_i = \lambda_i v_i$$

onde λ_i são os autovalores e v_i são os autovetores.

- 5) É criada a matriz P, com somente as colunas relativas aos componentes principais, identificados na etapa anterior.
- 6) A matriz de projeção, em função da matriz P, representa o número total de variáveis do processo em um espaço de dimensão reduzida em relação ao original.

As imagens a seguir ilustram resumidamente o procedimento em um espaço de 2 dimensões, utilizando a variável A e B para exemplificação. A Figura III-6: observações obtidas em um período normal do processo (dados de treinamento) das variáveis correlatas A e B, com elevado grau de associação. A Figura III-7: espaço das variáveis originais transformado, através da rotação dos eixos de coordenadas, em um espaço de variáveis latentes. O critério da máxima variância é utilizado para a determinação do novo eixo de coordenadas, que além disso deve ser ortonormal. A variável latente 1 pode ser considerada como o único componente principal, tendo em vista que ela claramente armazena quase a totalidade da variação dos dados, caracterizando esta reta como o subespaço principal (redução da dimensionalidade). Nesse contexto, a variável latente 2 seria uma variável residual (subespaço residual). A Figura III-8: novo dado do processo, projetado no subespaço principal (eixo da variável latente 1), e depois projetado novamente no espaço das variáveis originais. Desta forma, são obtidas as variáveis do processo, sem o impacto de resíduos/erros.

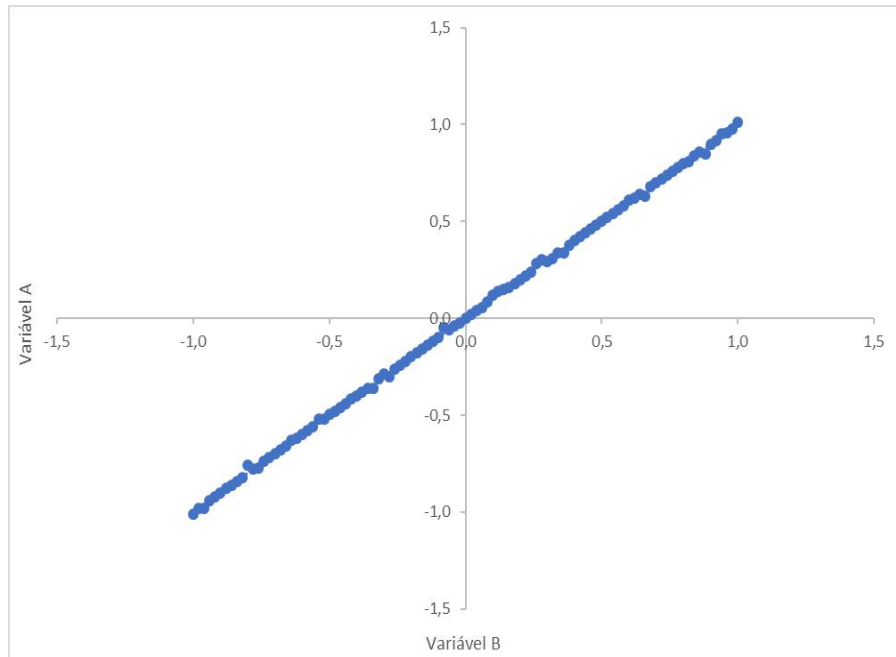


Figura III-6 Variáveis A e B plotadas a partir de uma base de dados de treinamento.

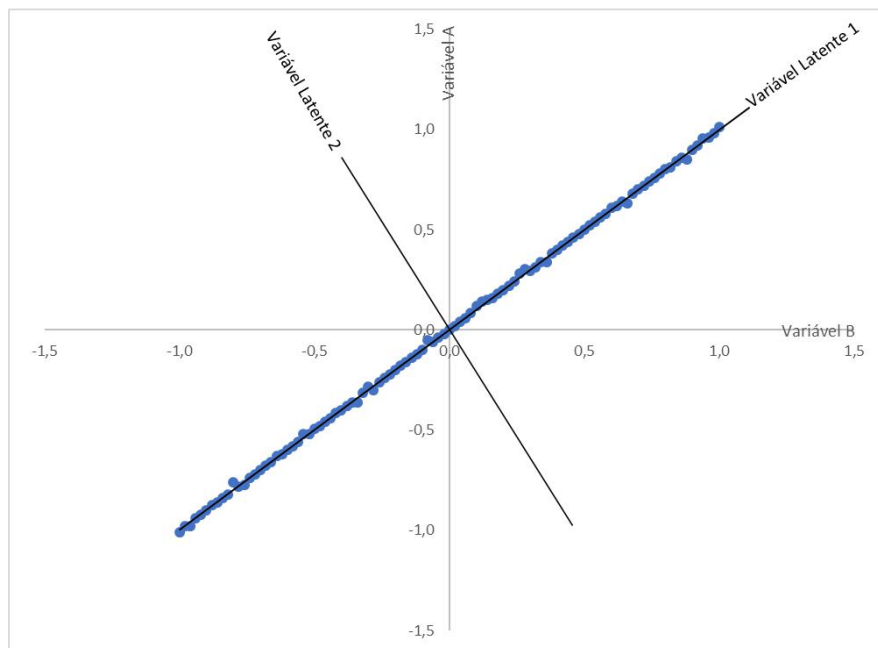


Figura III-7 Novos eixos de coordenadas (variáveis latentes), obtidos através do critério de máxima variância, além dos eixos de coordenadas originais.

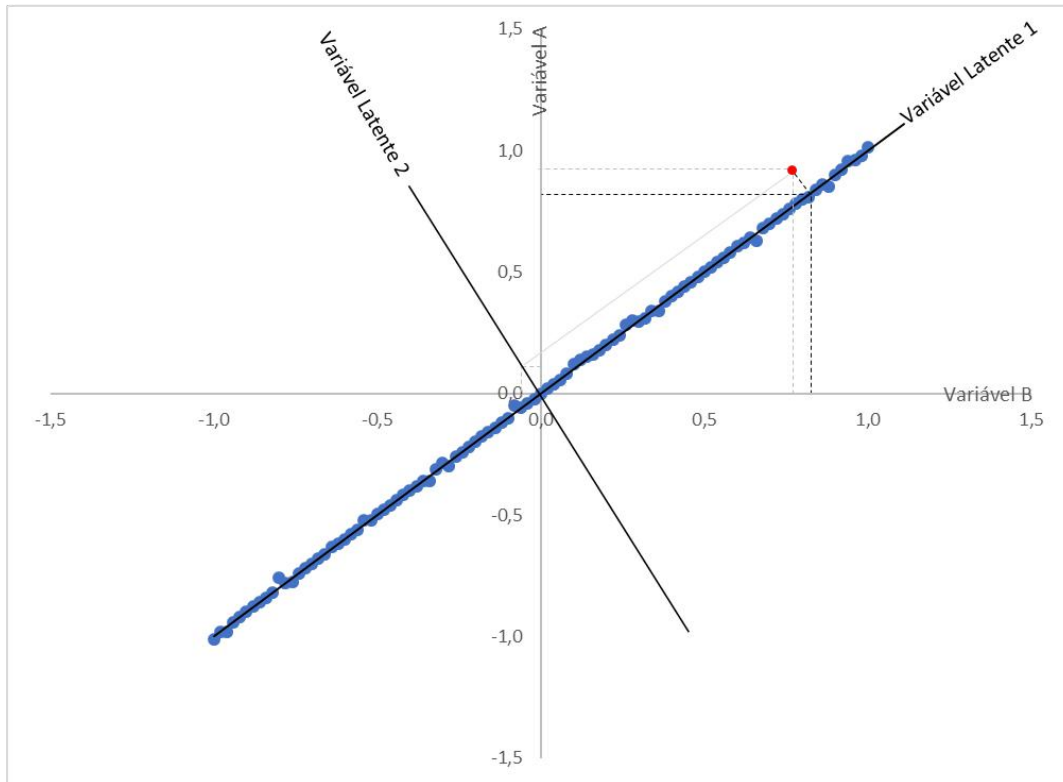


Figura III-8 Projeção de dado teste no subespaço principal, bem como posterior projeção no espaço original.

É importante observar que a técnica de PCA possui algumas limitações, de forma que sua aplicação para a resolução do problema apresentado possa demonstrar alguns inconvenientes. Dentre as limitações mais importantes para a avaliação dos resultados da aplicação da técnica para a resolução do problema apresentado neste trabalho estão: admitir que a relação entre as variáveis é linear; admitir que a ordem na sequência das observações é irrelevante; admitir que os dados representam o comportamento normal do processo (dentre o conjunto de dados de treinamento podem haver pontos de falha no processo. Com isso, o modelo poderia ser gerado considerando as falhas como variações normais do processo ou das vizinhanças); apresentar comportamento homocedástico (associação entre as variáveis é constante ao longo de todos os dados) e não multimodais (se os dados forem multimodais, haverá mais de uma tendência local na curva de probabilidade dos dados, ou seja, o processo terá pontos de operação diferentes) (Feital, 2011).

Dentre as limitações descritas, há uma que deve apresentar influência no modelo obtido por esse trabalho: a consideração de que os dados não são multimodais. Isto porque o

processo aqui estudado ocorre basicamente em dois pontos de operação, não respeitando essa premissa. Sobre o comportamento normal do processo e a ordem das observações, foi utilizada uma base de dados com conhecimento prévio de normalidade e em ordem cronológica. E sobre a associação linear entre as variáveis, foi demonstrado no início do trabalho que essa hipótese é satisfeita.

Para a detecção e o diagnóstico de falhas, foi utilizada a estatística Q, aqui representada pelo termo SPE (*Squared Prediction Error*), e foi realizada uma análise de contribuições (Qin, 2003).

O índice SPE mede a projeção do vetor da amostra no subespaço residual. O limite superior do SPE utilizado nesse trabalho está de acordo com o proposto por Qin (2003). A detecção de falha ocorre quando o SPE ultrapassa o limite superior calculado.

Após a detecção, deseja-se realizar o diagnóstico da falha. Gráficos de contribuição são a abordagem mais popular para o diagnóstico de falhas. Eles indicam a significância do efeito de cada variável no SPE ao longo do tempo. Se uma amostra x tiver uma SPE acima do limite, as variáveis que parecem ter uma contribuição significativa são investigadas. A partir desta análise, as variáveis com a maior contribuição são consideradas as principais colaboradoras para a falha. Para gerar os gráficos de contribuição não é necessário conhecimento prévio do processo. No entanto, o conhecimento prévio é frequentemente necessário para interpretar as parcelas de contribuição (Qin, 2003).

Se a redundância ou correlação entre as variáveis analisadas forem limitadas, é possível que algumas falhas não sejam identificáveis. Neste cenário, conclusões baseadas em qualquer método de diagnóstico devem ser cuidadosamente avaliadas (Qin, 2003).

CAPÍTULO IV - ANÁLISE DOS DADOS / DISCUSSÃO

IV.1 Visualização dos Dados

A seguir são apresentados os dados de vazão, em tonelada por hora, coletados hora-a-hora ao longo de cerca de um ano para três medidores redundantes, por onde passam a mesma corrente. Como a corrente não se divide e não se soma a outra no trajeto entre os medidores, espera-se que os instrumentos contabilizem aproximadamente a mesma quantidade de massa, conforme visualizado na Figura I-1, ressaltando-se que o Medidor virtual B é o somatório dos Medidores B1 e B2.

A vazão de distribuição de eteno ocorre em patamares, a depender da operação do cliente, que possui duas unidades de produção. Em patamares de fornecimento mais elevados, as duas unidades de produção estão em operação, enquanto o patamar mais baixo corresponde à vazão de fornecimento para a operação de somente uma das unidades. Além disso, são observados determinados períodos em que a vazão é zerada, isto ocorre em decorrência de paradas, programadas ou não, de produção. As Figuras IV-1 e IV-2 representam o perfil da vazão ao longo de um ano, onde ficam evidentes os comportamentos explicados.

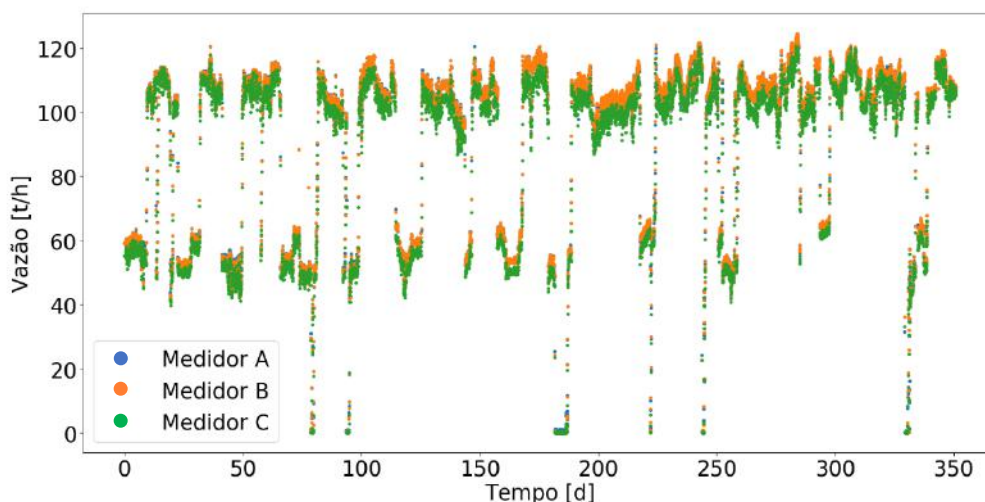


Figura IV-1 Vazão de produção de eteno, em toneladas por hora, ao longo de cerca de um ano.

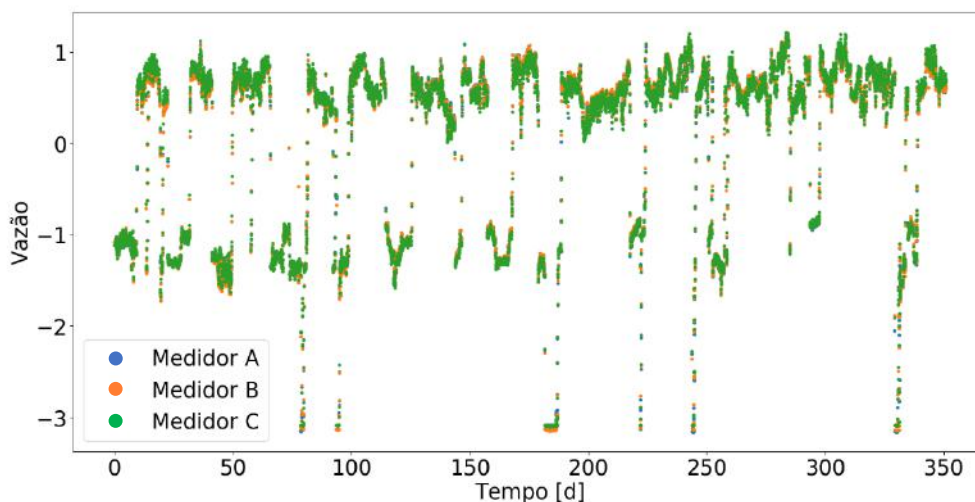


Figura IV-2 Vazão de produção de eteno, centralizada na média, ao longo de cerca de um ano.

IV.2 Caracterização dos Dados

IV.2.1 Correlação entre os medidores A e B

Ao se analisar a Figura IV-3, cujo eixo horizontal corresponde à vazão do Medidor A e o eixo vertical à vazão do medidor B, constata-se a alta correlação linear entre as vazões medidas pelos medidores A e B, tendo em vista que os pontos, relativos à vazão, se sobrepõem à reta de inclinação de 45° . Além disso, não é percebida uma diferença de correlação entre os medidores para diferentes patamares de medidas (Figuras IV-4 e IV-5).

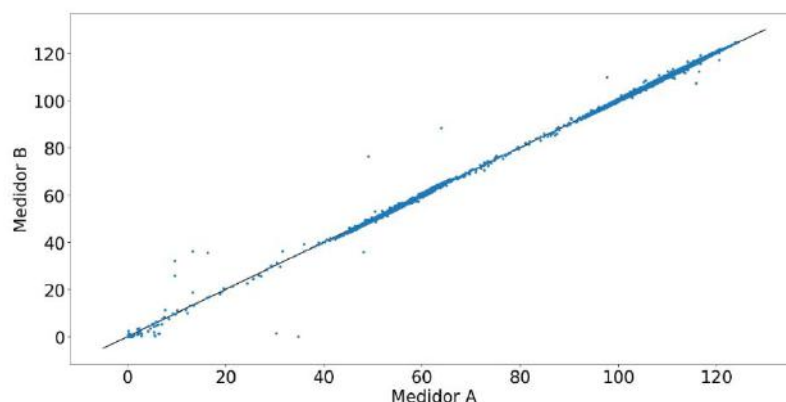


Figura IV-3 Gráfico de correlação entre as vazões, em toneladas por hora, dos Medidores A e B.

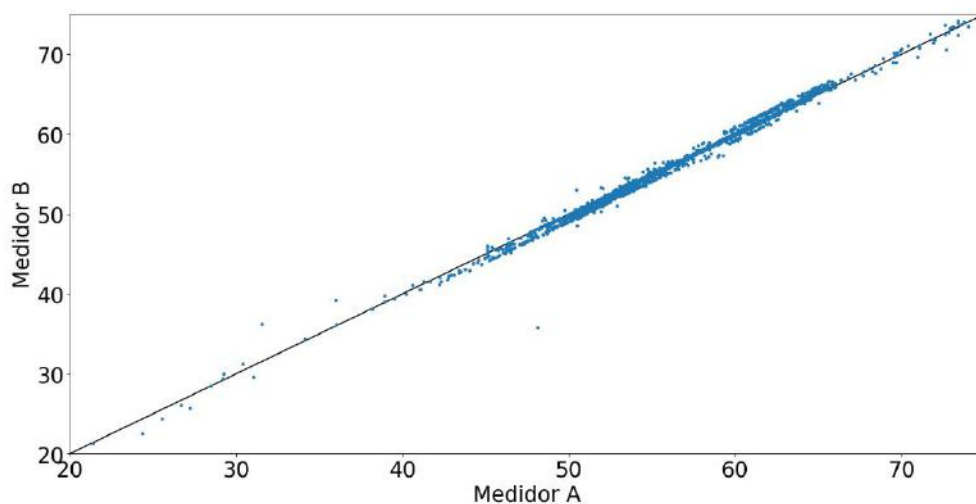


Figura IV-4 Gráfico de correlação entre as vazões [t/h] em patamares mais baixos dos Medidores A e B.

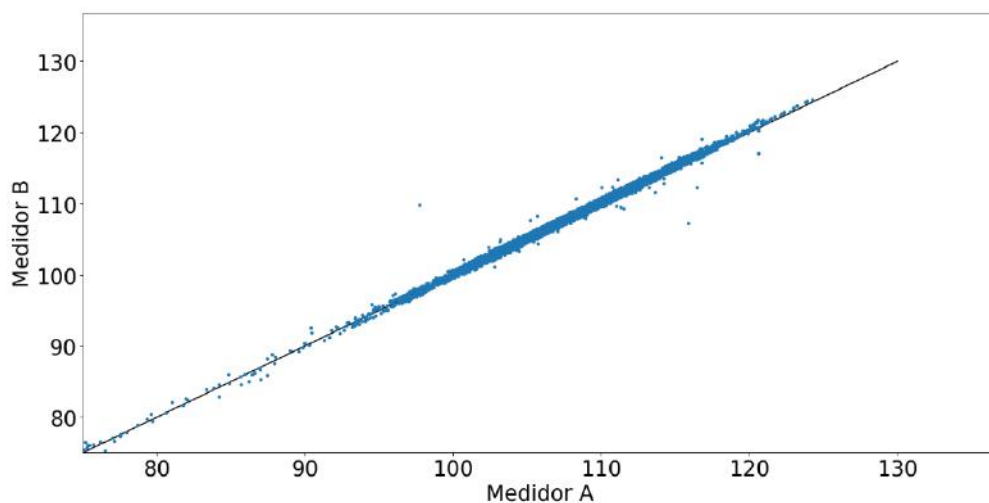


Figura IV-5 Gráfico de correlação entre as vazões [t/h] em patamares mais elevados dos Medidores A e B.

IV.2.2 Correlação do medidor C com os outros medidores

Percebe-se pelas Figuras IV-6, IV-7 e IV-8, que o Medidor C possui um desvio em relação aos outros dois instrumentos; tipicamente, este mede um valor de vazão abaixo dos outros e, aparentemente, esse viés se mantém ao longo de todo o patamar de vazão. No entanto, constata-se que o medidor C possui uma alta correlação linear com os outros dois

medidores, pois os pontos, relativos à vazão, ainda assim possuem uma tendência de inclinação de aproximadamente 45°.

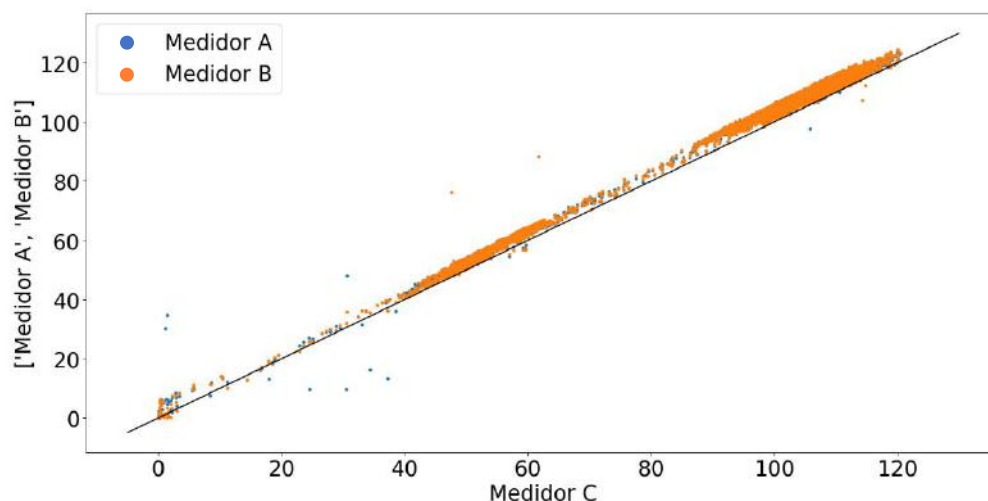


Figura IV-6 Gráfico de correlação entre a vazão, em toneladas por hora, do Medidor C com as vazões dos Medidores A e B, na mesma unidade de medida.

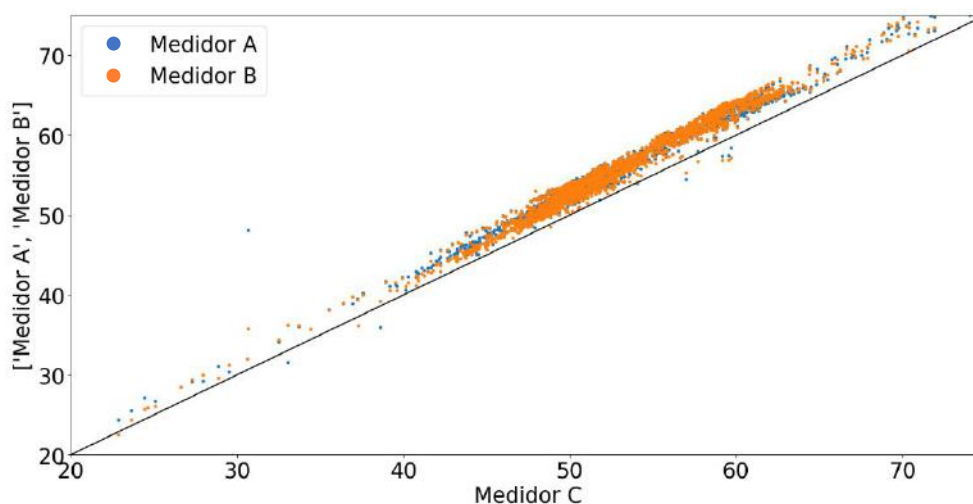


Figura IV-7 Gráfico de correlação entre a vazão em patamares mais baixos, em t/h, do Medidor C com as vazões dos Medidores A e B, na mesma unidade de medida.

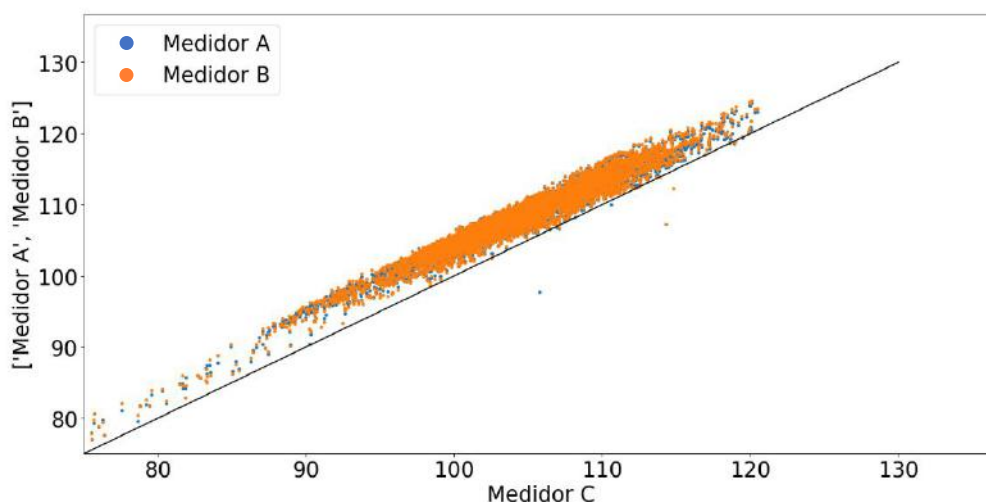


Figura IV-8 Gráfico de correlação entre a vazão em patamares mais elevados, em t/h, do Medidor C com as vazões dos medidores A e B, na mesma unidade de medida.

IV.3 Espectro de Variância

Os espectros de variâncias aqui apresentados foram construídos utilizando-se bases de dados montadas de formas diferentes para as diferentes análises, como detalhado abaixo. Além disso, foram construídos espectros utilizando-se tanto a mediana como a média como medida de tendência central. Os espectros gerados das duas formas foram muito similares, o que é uma indicação de ausência de valores discrepantes no conjunto de dados. Por isso, neste trabalho serão apresentados somente os espectros elaborados empregando a média como medida de tendência central dos dados.

IV.3.1 Análise com comprimento de janela grande

Para a análise da variância relativa a grandes comprimentos de janela, foram utilizados os dados capturados hora-a-hora ao longo de cerca de um ano de processo, de forma a ser necessário um tempo razoável para o processamento dos dados. Nesse caso, mesmo que a utilização de dados captados de hora em hora omita boa parte da variação da vazão através do emprego da média da vazão nesses períodos, como o objetivo não é a análise de erros de medição de alta frequência, foi possível estruturá-los desta forma.

Nos espectros aqui presentes, o eixo vertical corresponde ao desvio padrão da vazão (em toneladas por hora) para os tamanhos de janelas (em dias), mostrados no eixo horizontal.

Os espectros abaixo mostram que o desvio padrão total para a amostra analisada é cerca de 17,55 t/h, 17,62 t/h e 17,27 t/h para os medidores A, B e C, respectivamente (Figura IV-9).

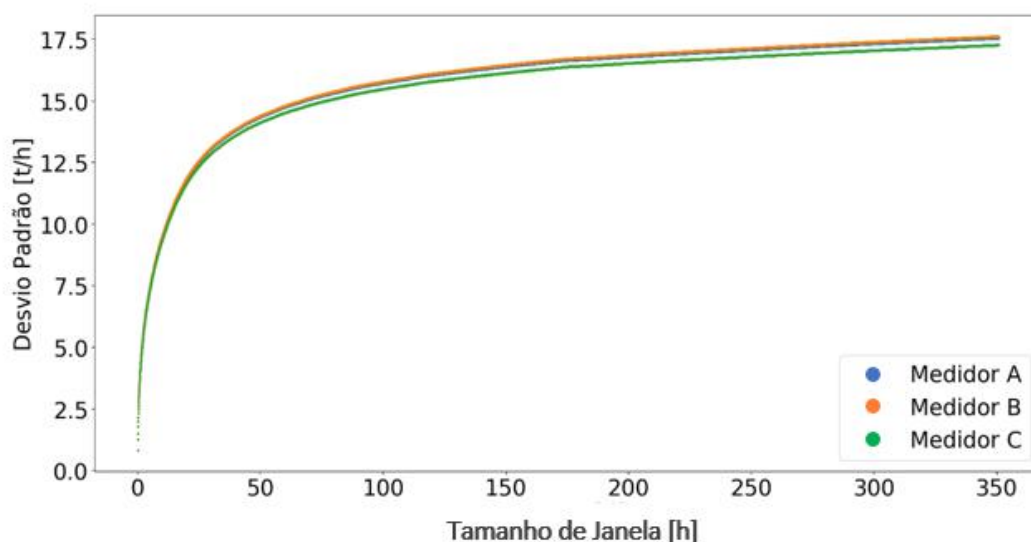


Figura IV-9 Espectro de Variância dos Medidores A, B e C, utilizando-se dados de vazão hora-a-hora ao longo de cerca de um ano.

Além disso, pode-se dizer que, para o ano analisado, quase toda a dinâmica do processo pôde ser observada durante períodos de cerca de 44 dias, correspondente a aproximadamente 80% do desvio padrão total observado para o conjunto de dados, conforme Figura IV-10.

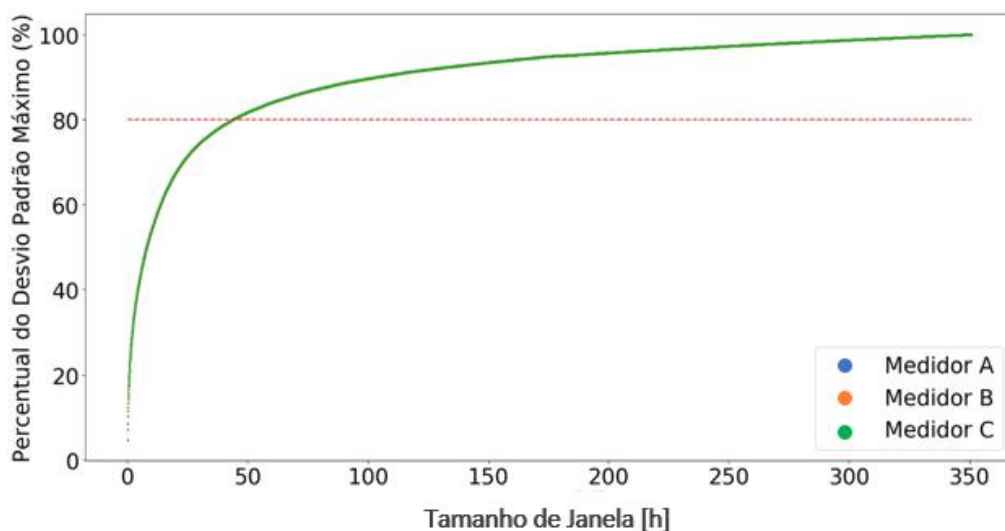


Figura IV-10 Percentuais, em relação ao desvio padrão máximo da amostra analisada, para os tamanhos de janelas percorridos em dias.

Para fins de comparação mais à frente neste trabalho, também foram calculados os desvios padrão dos medidores para tamanho de janela igual a 2, são eles: 0,81 t/h, 0,81 t/h e 0,79 t/h, correspondentes aos medidores A, B e C, respectivamente (Figura IV-11).

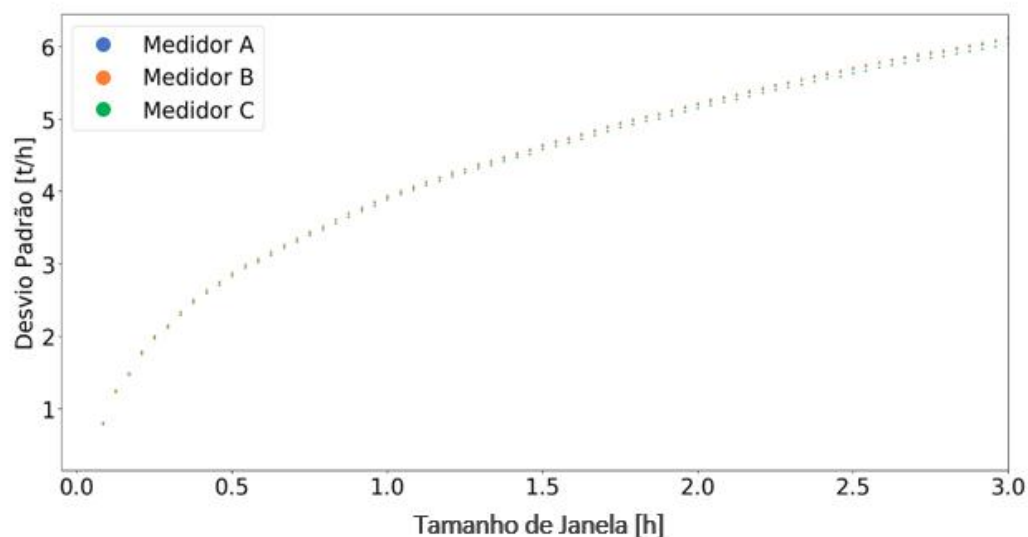


Figura IV-11 Espectro de Variância dos Medidores A, B e C, utilizando-se dados de vazão hora-a-hora, com o foco em janelas de pequenos tamanhos.

IV.3.2 Análise com comprimento de janela pequeno

Para a análise da variância relativa à pequenos comprimentos de janela, foram utilizados os dados capturados minuto-a-minuto ao longo de quatro dias que englobavam os diferentes patamares de operação, de forma a ser necessário um tempo razoável para o processamento dos dados. Nesse caso, como o objetivo é a análise de erros de medição de alta frequência, foram utilizados os dados estruturados desta forma, a fim de não mascarar variações de alta frequência com o uso da média.

Os desvios padrão dos medidores para tamanho de janela igual a dois foram: 0,0030 t/min, 0,0020 t/min e 0,0022 t/min, correspondentes aos medidores A, B e C, respectivamente (Figuras IV-12 e IV-13). A Tabela IV.1 apresenta esses valores transformados para a unidade de grandeza tonelada por hora, a fim de comparar com os desvios padrão obtidos com os dados hora-a-hora do ano inteiro.

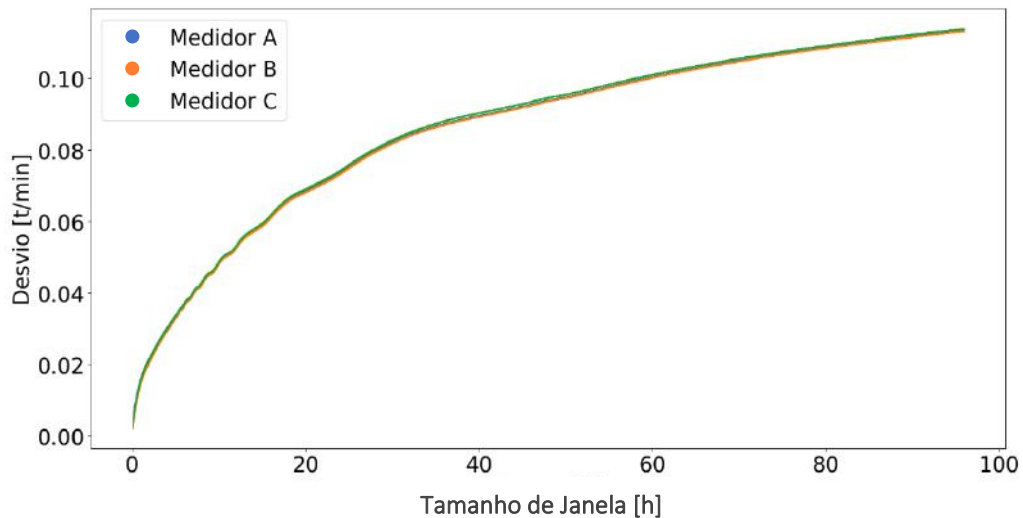


Figura IV-12 Espectro de Variância dos Medidores A, B e C, utilizando-se dados de vazão minuto-a-minuto ao longo de quatro dias.

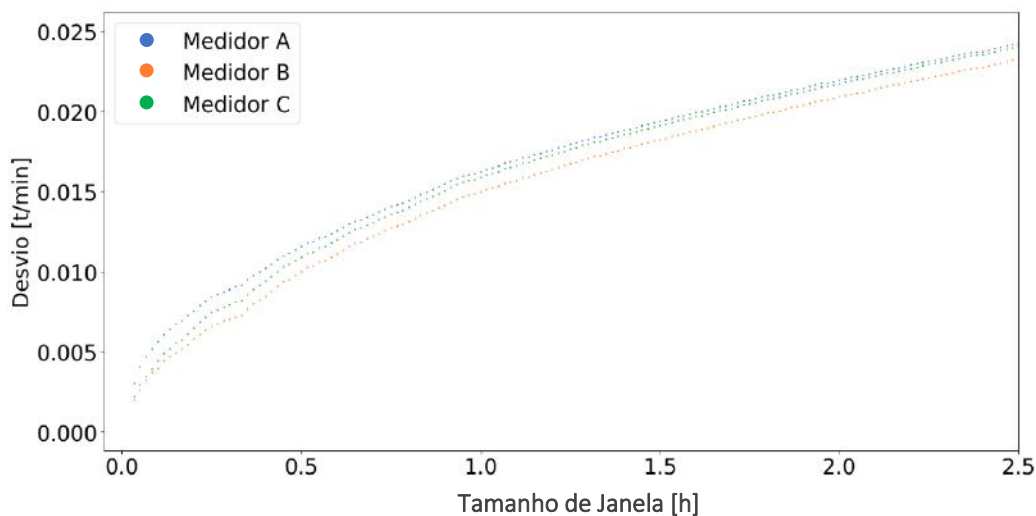


Figura IV-13 Espectro de Variância dos Medidores A, B e C, utilizando-se dados de vazão minuto-a-minuto, com o foco em janelas de pequenos tamanhos.

Tabela IV-1 Desvios padrão, em toneladas por hora, observados com tamanho de janela igual a 2 para os Medidores A, B e C, utilizando bases de dados estruturadas de formas diferentes.

Medidores\Desvio padrão	Base minuto-a minuto	Base hora-a-hora
A	0,18 ± 0,01	0,81 ± 0,04
B	0,12 ± 0,01	0,81 ± 0,04
C	0,13 ± 0,01	0,79 ± 0,04

Fonte: própria (2020)

Como já previamente discutido, resumidamente, em janelas de tamanho pequenos, mais especificamente de tamanho 2, espera-se capturar ruídos e outros erros de medição. A Tabela IV-1 apresenta os desvios padrão das medições para uma janela de tamanho 2, isso quer dizer que os erros de alta frequência podem ter um impacto na medição, para mais ou para menos, de potencialmente, no máximo, os valores apresentados na tabela. Além disso, de forma a realizar uma comparação estatística entre os desvios padrão apresentados, foram calculados os erros padrão.

Ao realizar a comparação entre o desvio padrão observado ao se basear no conjunto de dados minuto-a-minuto e ao ser utilizado o conjunto hora-a-hora, constata-se que, para a primeira forma de armazenar os dados, o desvio padrão é menor. Isso pode ser explicado pelo fato de que, apesar de a média empregada para armazenagem do dado de vazão no período de 1 hora de certa forma amortecer parte de variações de alta frequência, que ocorreriam nesse período de tempo, ao mesmo tempo, fazendo isso, os dados passam a operar em patamares, cujas subidas e descidas das médias de hora em hora impactam no cálculo da variância calculada para pequenas janelas (que capturam a mudança potencialmente abrupta dos dados de vazão de uma hora de acompanhamento para a outra).

Os dados apresentados na tabela evidenciam que os Medidores B e C apresentam, estatisticamente (intervalo de confiança de 95%), o mesmo desvio padrão, o que implica na conclusão de que as medições indicadas por eles são impactadas na mesma magnitude por erros de alta frequência, mesmo que a qualificação desses erros seja diferente para cada um dos três medidores. Já o Medidor A possui um desvio padrão maior que os outros, mas somente quando calculado com a base de dados minuto-a-minuto.

Considerando-se que, coerentemente, o desvio padrão obtido para o tamanho de janela dois seja simplesmente um erro de medição. Para o cenário mais extremo possível, isso refletiria em uma aferição da transferência erroneamente, considerando-se o medidor A, por exemplo, de cerca de 0,18t/h de forma constante; ao longo de um ano, essa quantidade se totalizaria em cerca de 1.577 toneladas. Nesse caso, essa quantidade a mais estaria sendo transferida incorretamente, tendo em vista que pode ser considerada um erro. Economicamente, considerando uma margem de cerca de R\$212/tonelada de eteno (Mello,

2017), isso representa um impacto de cerca de R\$334.000 no ano, neste caso favorecendo a produtora de eteno e desfavorecendo o recebedor do eteno (planta de polietileno).

Além de movimentos de transferência entre plantas da mesma empresa, muitas indústrias realizam vendas para clientes externos e compras de fornecedores. Para um processo contínuo, boa parte dessas quantidades também são consolidadas através das indicações de medidores de vazão. Então, o impacto observado para o eteno neste trabalho deve ser entendido como um indicativo de que outros instrumentos devem ser analisados a fim de qualificá-los, tanto por uma questão de *compliance*, quanto por uma questão econômica, pois, diferentemente para o que foi constatado para a corrente de eteno, outras correntes, de matérias-primas, por exemplo, podem estar acarretando em uma perda financeira.

IV.4 Análise de Componentes Principais

Para a realização da análise de componentes principais, foi utilizada a base de dados coletada hora-a-hora. Os dados foram divididos em dois grupos: no primeiro grupo foi alocada a maior quantidade de dados (cerca de 10 meses), já que este grupo seria o conjunto para treinamento da técnica de PCA, ou seja, é com esse conjunto de dados que os parâmetros estatísticos que servem como base para a técnica seriam originados. O segundo conjunto de dados foi utilizado para fim de teste da técnica, ou seja, sabendo-se que esses são dados de operação normal do processo, eles foram empregados para verificar se não seriam gerados falso-positivos logo após a implementação da técnica para fins de monitoramento e detecção de erros.

IV.4.1 Percentagem de variância cumulativa

Suportado pela observação da Figura IV-14, pode-se dizer que a primeira variável latente armazena 99,94% da variabilidade dos dados, portanto, é nela que está contida a maior parte da informação que impacta nas medições observadas. A segunda variável latente, 0,04% e a terceira, 0,02%, como pode ser visualizado através da ampliação dada aos dados na Figura IV-15. Com isso, pode-se dizer que há somente uma variável principal que representa o comportamento do conjunto de dados; essa variável principal seria a variável latente 1. De fato, era esperado que esse comportamento fosse identificado, pois essa variável

representaria a vazão real do processo, que é o sinal intrínseco que modela o comportamento das variáveis medidas, sem a influência de ruídos e demais erros, que estariam contidos nas variáveis latentes 2 e 3.

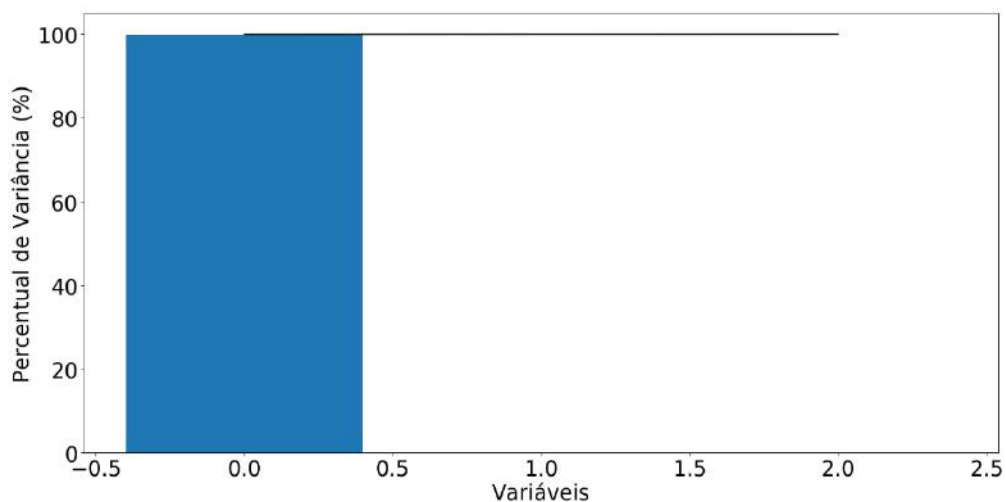


Figura IV-14 Percentagem de variância cumulativa das três variáveis latentes reveladas pela PCA.

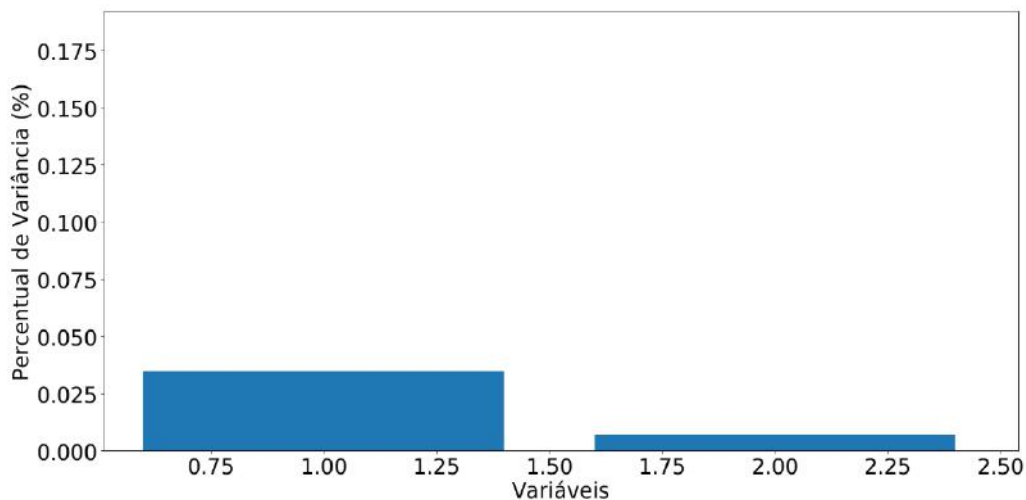


Figura IV-15 Percentagem de variância cumulativa das variáveis latentes, não caracterizadas como variáveis principais, conforme reveladas pela PCA.

IV.4.2 Variáveis originais e Variáveis latentes

Os gráficos a seguir apresentam as variáveis observadas no lado esquerdo e as variáveis latentes no lado direito. O eixo horizontal corresponde à quantidade de dias corridos. A

separação por cores a partir de determinado dia, demarca a parte gráfica utilizada como treino e a parte utilizada como teste.

Pelos gráficos das Figuras IV-16 e IV-17, é perceptível que de fato a primeira variável latente é a única variável principal, notando-se que o comportamento desta está intimamente ligado ao comportamento das três variáveis observadas (lado esquerdo). Nesse contexto, a segunda e terceira variáveis latentes seriam variáveis resíduos, que impactariam imputando uma quantidade de erro nas variáveis observadas. A fim de se obter o valor da vazão verdadeiro, foram calculadas novas vazões descontando-se as contribuições das variáveis residuais. Por isso, os gráficos à esquerda apresentam duas variáveis cada, que praticamente se sobrepõem ao longo de todo o tempo; uma é a variável observada e a outra é a variável verdadeira, ou seja, sem a contribuição dos erros (resíduos).

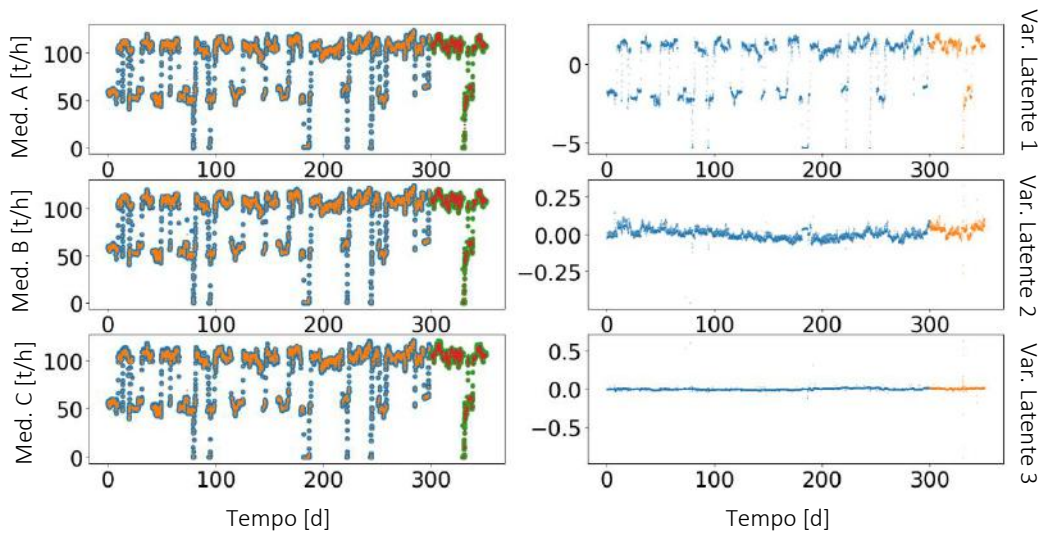


Figura IV-16 Visão inicial ao longo do tempo (dias) das três variáveis observadas (lado esquerdo) – em t/h – e das três variáveis latentes (lado direito).

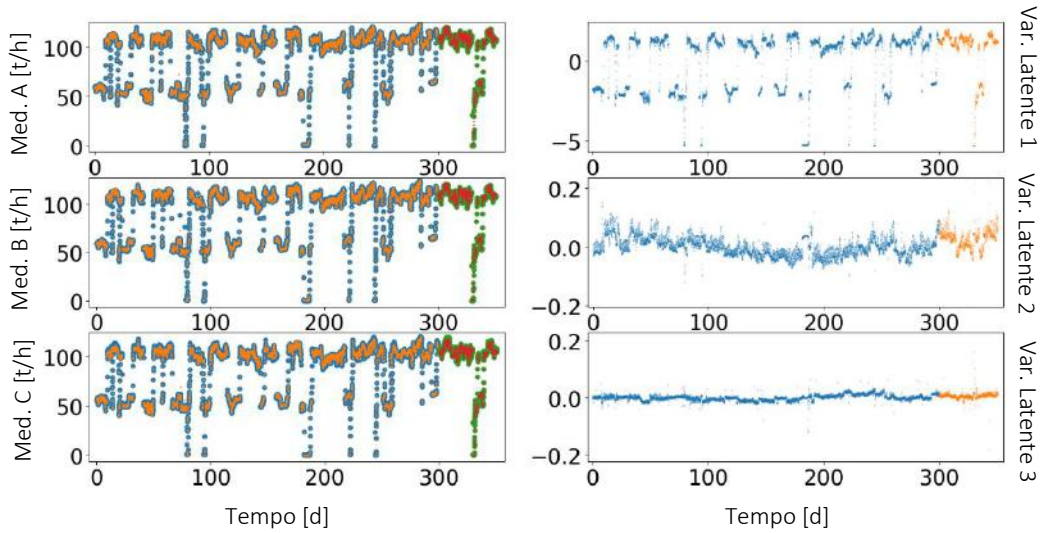


Figura IV-17 Visão ao longo do tempo (dias) das três variáveis observadas (lado esquerdo) – em t/h – e das três variáveis latentes (lado direito).

Ao analisar as variáveis em períodos de tempo mais curtos, pôde-se perceber um comportamento periódico na segunda variável latente (Figuras IV-18 e IV-19). Ao observar um período de 7 dias, claramente percebe-se que a variável latente 2 inicia o dia com determinado valor, que aumenta até o meio do dia, quando então volta a diminuir até o fim do dia; este comportamento é observado durante todos os dias, sejam dias onde constam os dados de treinamento ou os dias onde constam os dados de teste (Figura IV-20). A fim de se investigar a possível motivação para este comportamento, foram pesquisados dados de temperatura ambiente ao longo dos dias e percebe-se que o comportamento observado na variável latente é altamente correlacionado à variação da temperatura ambiente diária, tendo em vista que o perfil de variação diária e anual é o mesmo (Figuras IV-22 e IV-23). Além disso, parece que também a variável 2 possui uma variação, mesmo que bastante suave, ao longo dos meses; reduzindo a partir do dia 100 (abril) e voltando a aumentar aproximadamente no dia 250 (setembro), o que corresponde à variação de temperatura anual (Tabela IV-2 e Figura IV-21) .

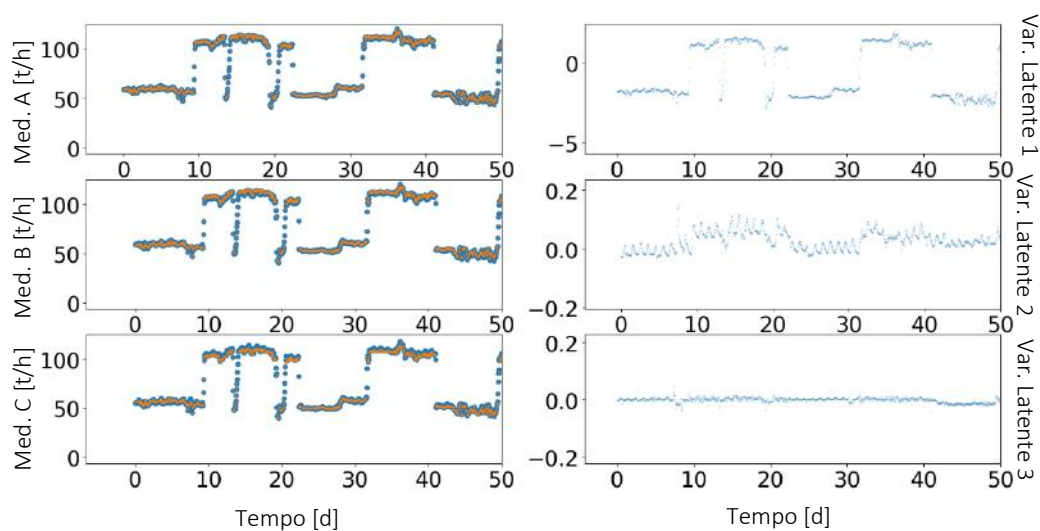


Figura IV-18 Visão ao longo do tempo das três variáveis observadas (lado esquerdo) – em t/h – e das três variáveis latentes (lado direito) no período de 50 dias corridos.

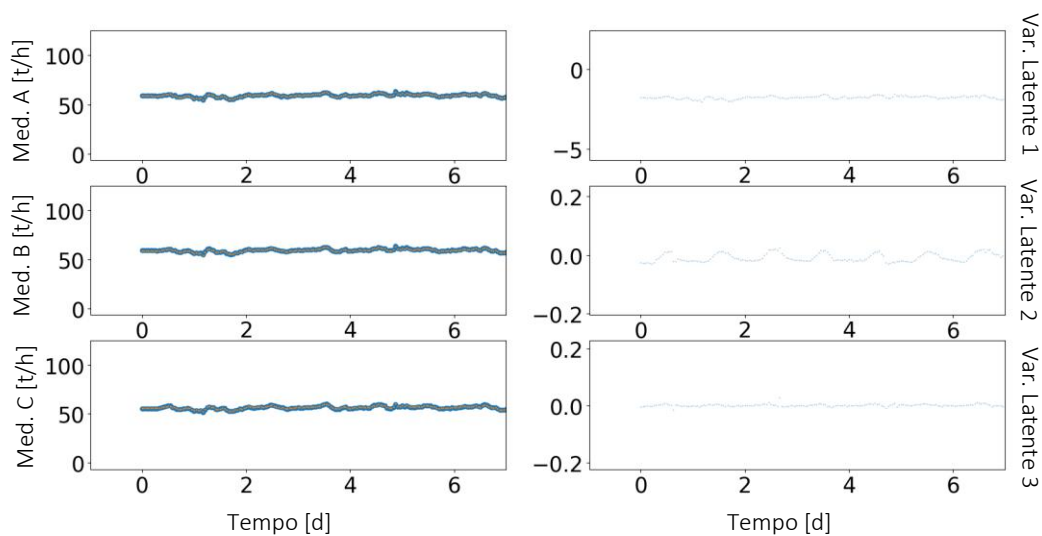


Figura IV-19 Visão ao longo do tempo das três variáveis observadas (lado esquerdo) – em t/h – e das três variáveis latentes (lado direito) no período de 7 dias corridos.

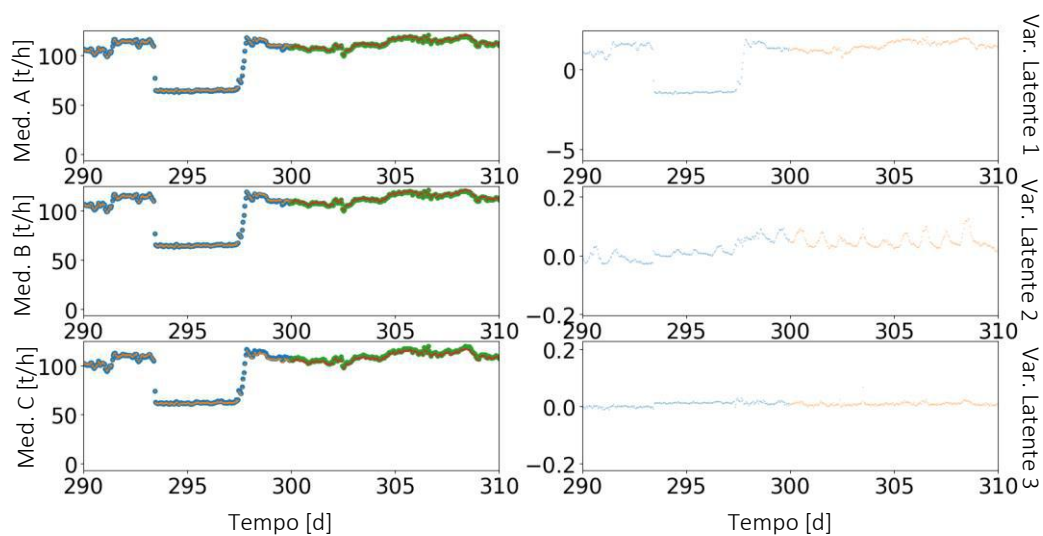


Figura IV-20 Visão ao longo do tempo (dias) das três variáveis observadas (lado esquerdo) – em t/h – e das três variáveis latentes (lado direito), com foco no período de transição entre os dados de treinamento e os dados de teste.

Tabela IV-2 Temperaturas médias típicas mensais na região metropolitana do Rio de Janeiro.

Mês	Minima (°C)	Máxima (°C)
Janeiro	22°	32°
Fevereiro	22°	33°
Março	21°	32°
Abril	19°	29°
Maio	17°	28°
Junho	15°	27°
Julho	15°	26°
Agosto	16°	28°
Setembro	17°	28°
Outubro	19°	28°
Novembro	20°	29°
Dezembro	21°	31°

Fonte: Climatempo (2020)

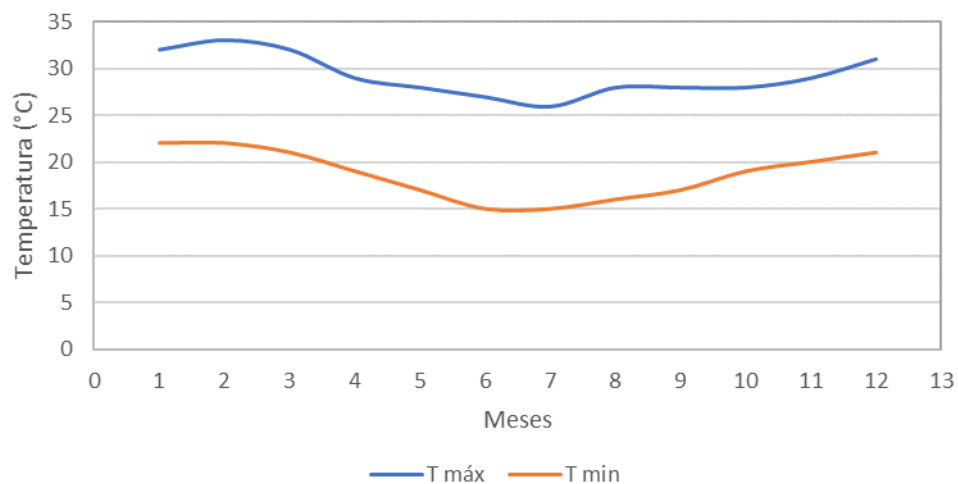


Figura IV-21 Perfil de temperatura (°C) ao longo do ano na região metropolitana do Rio de Janeiro. Construído através dos dados de Climatempo (2020).

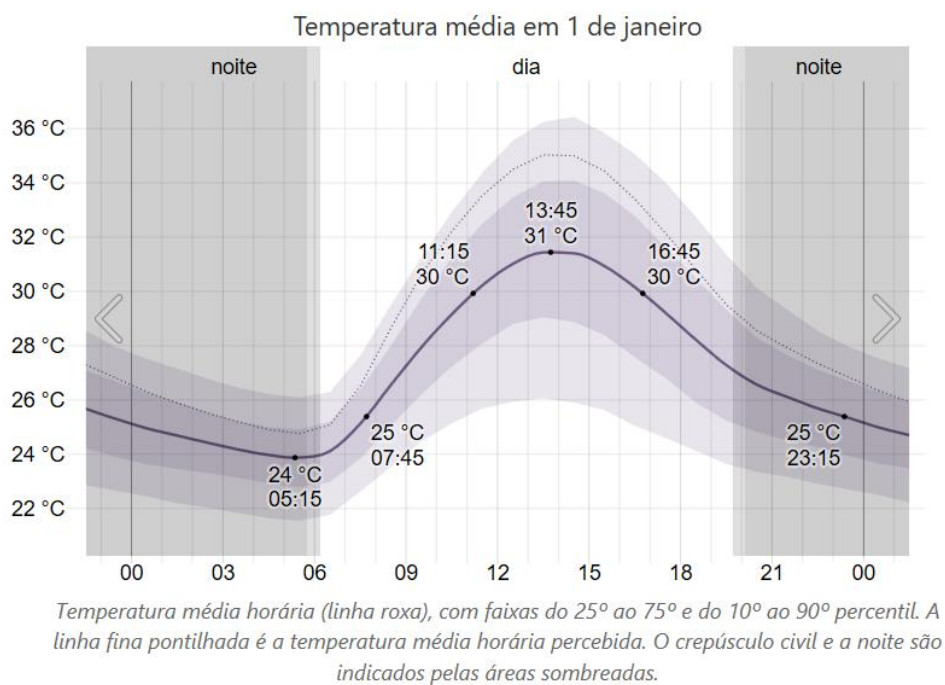


Figura IV-22 Perfil de temperatura média (°C) ao longo de 01 de janeiro.

Fonte: Weather Spark (2020)

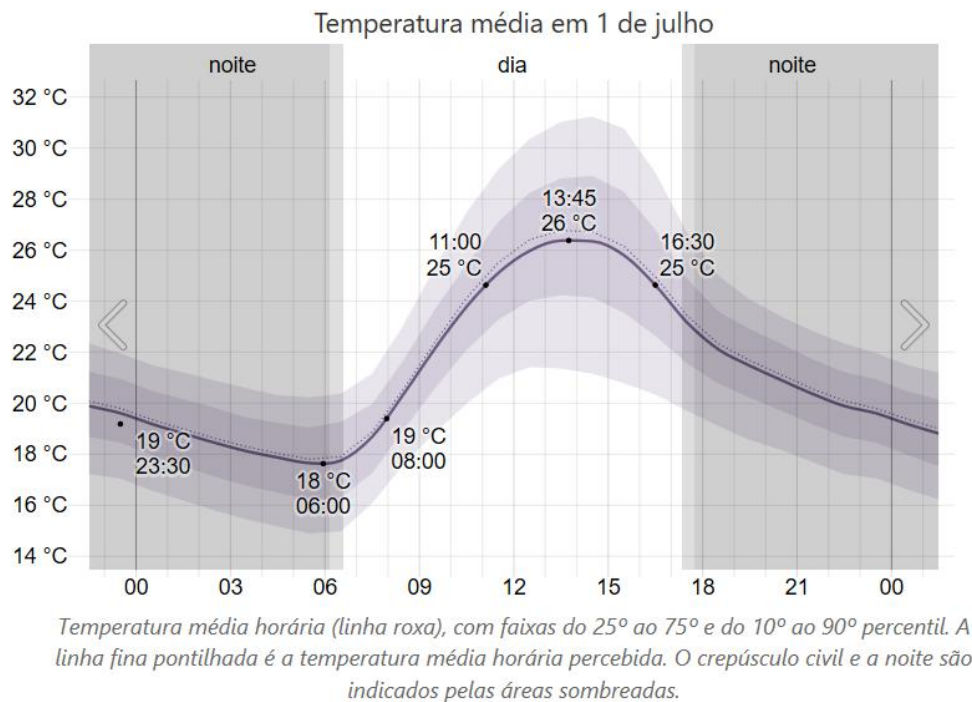


Figura IV-23 Perfil de temperatura média (°C) ao longo de 01 de julho.

Fonte: Weather Spark (2020)

A terceira variável latente ficou praticamente constante e próxima de zero ao longo de todo o tempo. Ela será discutida mais à frente, pois apresentou um comportamento importante ao ser simulado um erro de medição nos dados.

Ademais, ao se analisar o período de transição dentre os dados de treinamento e os dados de teste, pode-se constatar que os parâmetros estatísticos basais para a técnica foram gerados com uma base de dados adequada, tendo em vista que o comportamento das variáveis latentes não mudou abruptamente ao haver a mudança do período de treinamento para o período de teste (Figura IV-24).

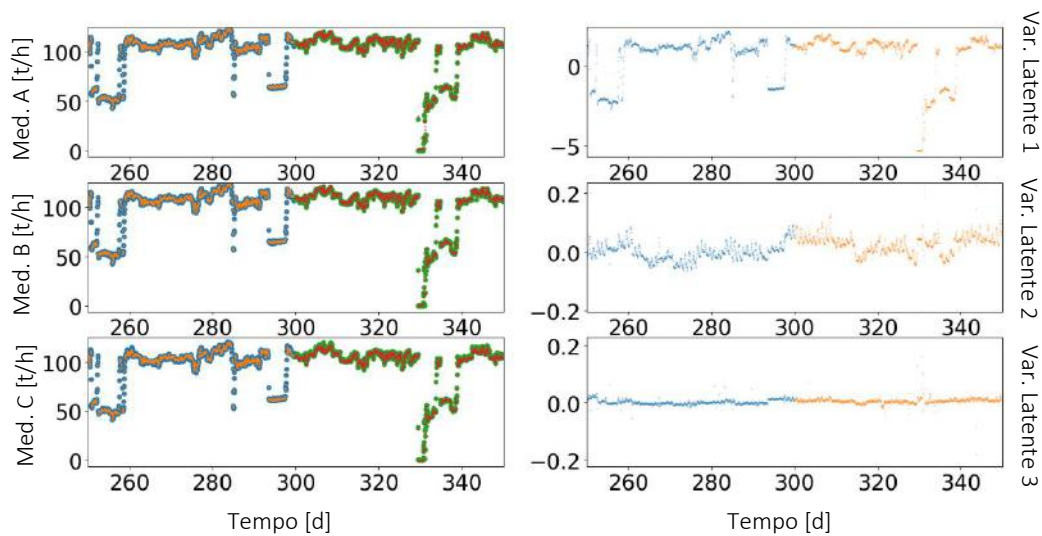


Figura IV-24 Visão ao longo do tempo (dias) das três variáveis observadas (lado esquerdo) – em t/h – e das três variáveis latentes (lado direito), com foco no período de transição entre os dados de treinamento e os dados de teste.

IV.4.3 Squared Prediction Error (SPE)

Conforme já comentado, para monitoramento a detecção de erro nas medições, foi utilizada a estatística Q, resultando no gráfico mostrado pela Figura IV-25. Conforme pode ser observado, em alguns períodos o limite é atingido com certa frequência (Figura IV-26). Para que seja realizado um diagnóstico *online*, seria importante que fossem determinadas premissas de tolerância de pontos fora do limite, similar, por exemplo, ao que é feito nas cartas de controle de Shewhart, para que a indicação de erro e necessidade de manutenção seja efetiva, minimizando os falsos-positivos.

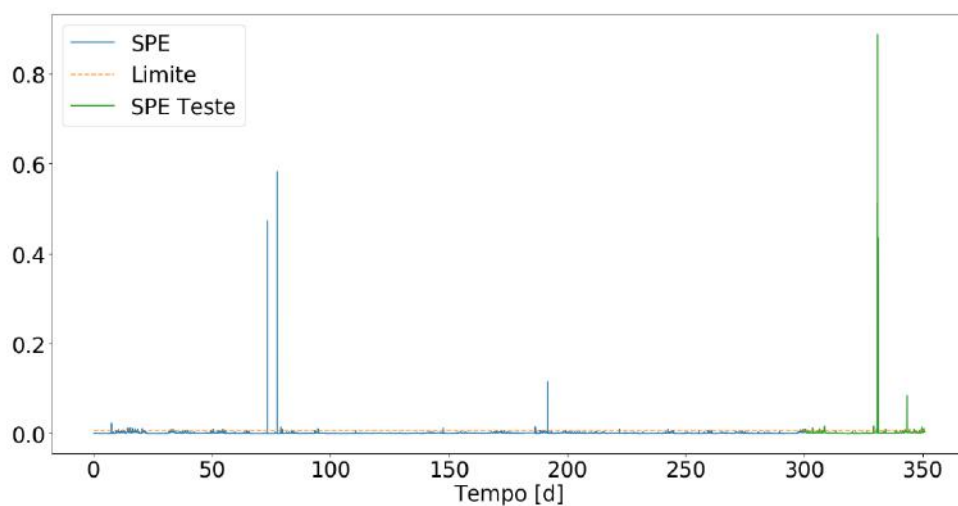


Figura IV-25 SPE e limite de erro calculado conforme a literatura (Qin, 2003).

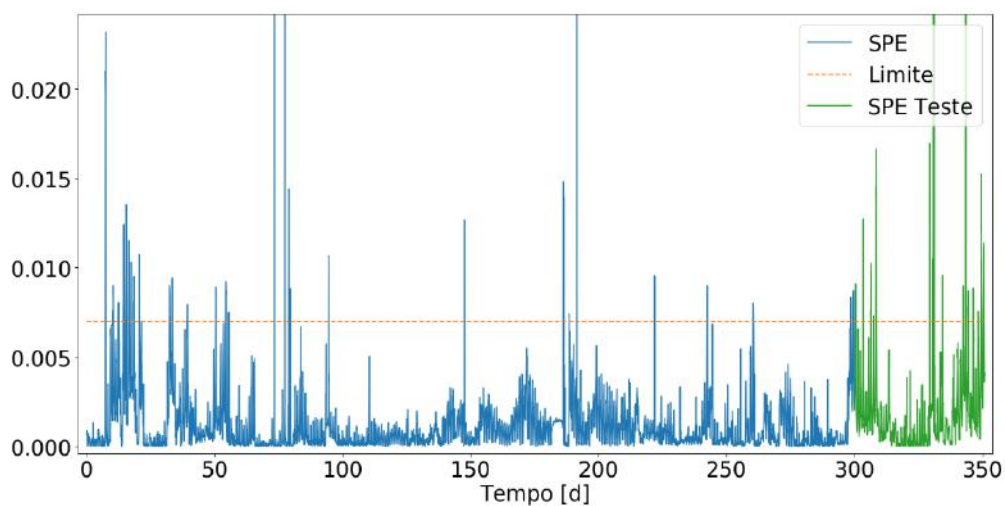


Figura IV-26 SPE e limite de erro calculado conforme a literatura (Qin, 2003).

Além disso, ao se analisar o período de transição dentre os dados de treinamento e os dados de teste, pode-se constatar que os parâmetros estatísticos basais para a técnica foram gerados com uma base de dados adequada, tendo em vista que o SPE continuou com um comportamento similar ao que estava ocorrendo e não ultrapassou o limite pré-determinado durante todo o tempo depois. (Figura IV-27)

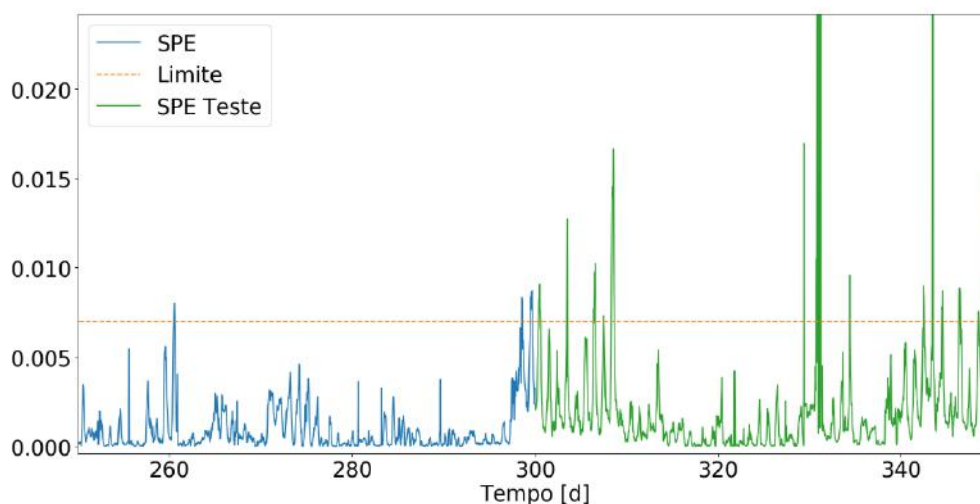


Figura IV-27 SPE e limite de erro calculado conforme a literatura (Qin, 2003), com foco no período de transição entre os dados de treinamento e os dados de teste.

IV.4.4 Contribuições

Após identificar o momento em que ocorre um erro com algum medidor, é necessário identificar em qual medidor está ocorrendo o erro. Para isso, deve-se analisar as contribuições de cada medidor no erro. Abaixo, são mostrados os gráficos de contribuições dos três medidores estudados, tanto para os dados de treinamento, quanto para os dados de teste; os conjuntos de dados estão separados por diferentes cores.

Observa-se que o medidor C normalmente é o que oferece maior contribuição ao erro, quase durante todo o tempo (Figuras IV-29 e IV-30). Como exemplo, é possível analisar, especificamente, o período inicial de 25 dias. Durante esse tempo, percebe-se que o SPE ultrapassa o limite com frequência (Figura IV-26); ao fazer a análise das contribuições, conclui-se que o medidor que mais contribuiu para a elevação desse erro durante esse tempo foi o Medidor C (Figuras IV-29 e IV-31). Outro exemplo seriam os últimos dias acompanhados, quando o SPE ultrapassa com frequência o limite; nesse momento, também se analisa que grande parte da contribuição para o erro decorre do Medidor C.

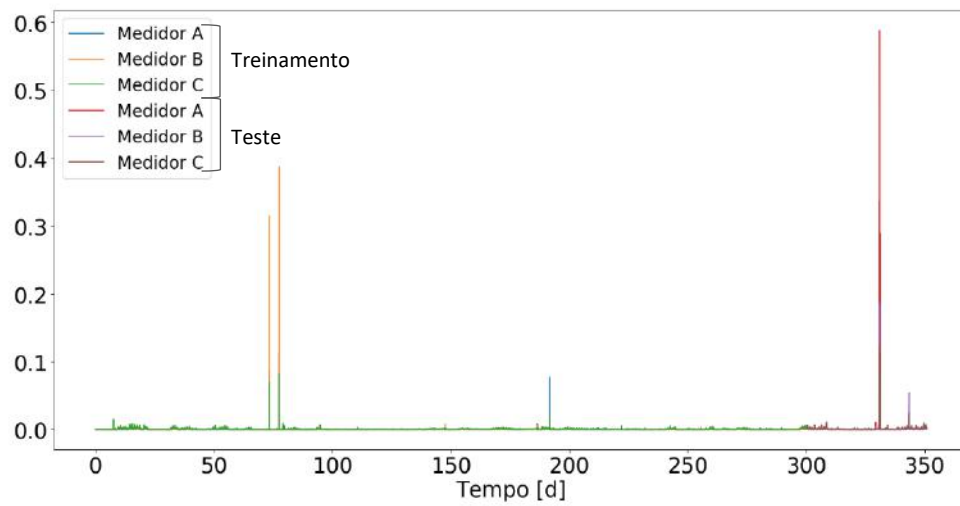


Figura IV-28 Visão inicial ao longo do tempo (dias) das contribuições dos três medidores no erro, utilizando-se dados para treinamento e dados para teste (a partir do dia 300).

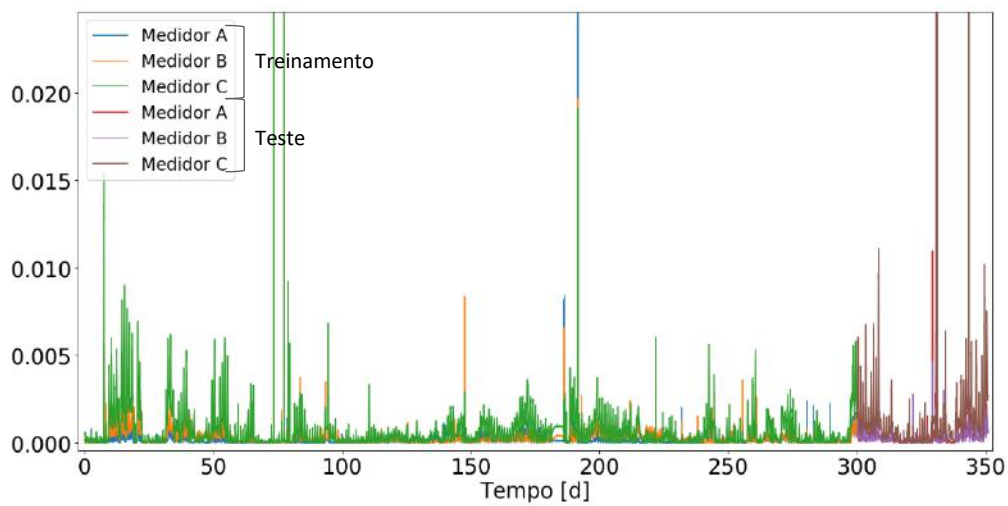


Figura IV-29 Visão ao longo do tempo (dias) das contribuições dos três medidores no erro, utilizando-se dados para treinamento e dados para teste (a partir do dia 300).

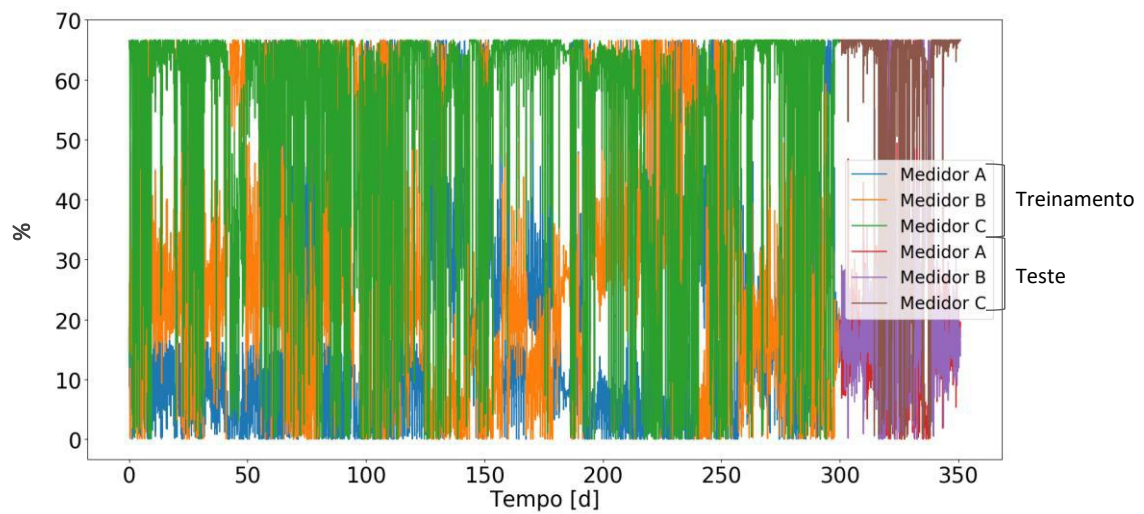


Figura IV-30 Visão ao longo do tempo (dias) do percentual de cada contribuição para o SPE, utilizando-se dados para treinamento e dados para teste (a partir do dia 300).

Apesar de normalmente a maior contribuição para o erro ser do medidor C, é interessante observar que mesmo que não ultrapasse o limite estipulado para o SPE, em alguns momentos esse perfil muda, por exemplo, entre os dias 42 e 50, a maior contribuição para o erro passa a ser do Medidor B (Figuras IV-31 e IV-32).

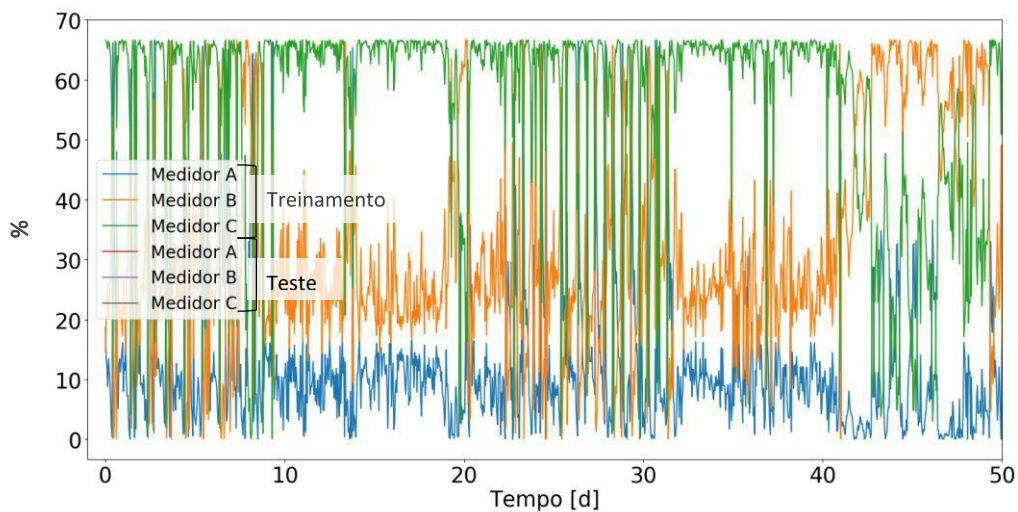


Figura IV-31 Visão ao longo do tempo (dias) do percentual de cada contribuição para o SPE, no período de 50 dias corridos.

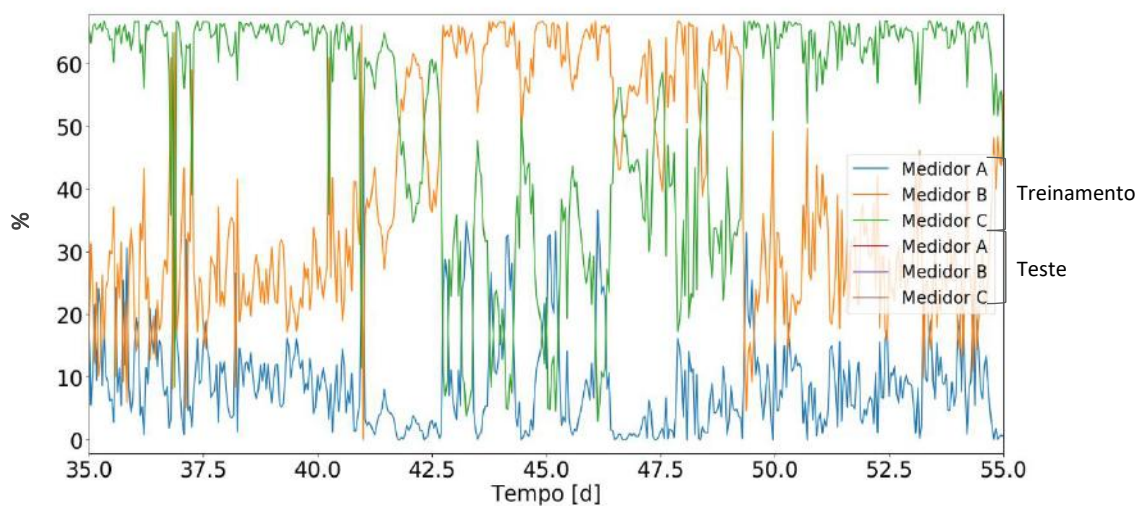


Figura IV-32 Visão do percentual de cada contribuição para o SPE, entre os dias 35 e 55.

Além disso, em relação à transição para os dados de teste, observa-se que o comportamento das contribuições segue com um perfil similar ao que tinha com a aplicação nos dados de treinamento, para os três medidores (Figuras IV-33 e IV-34).

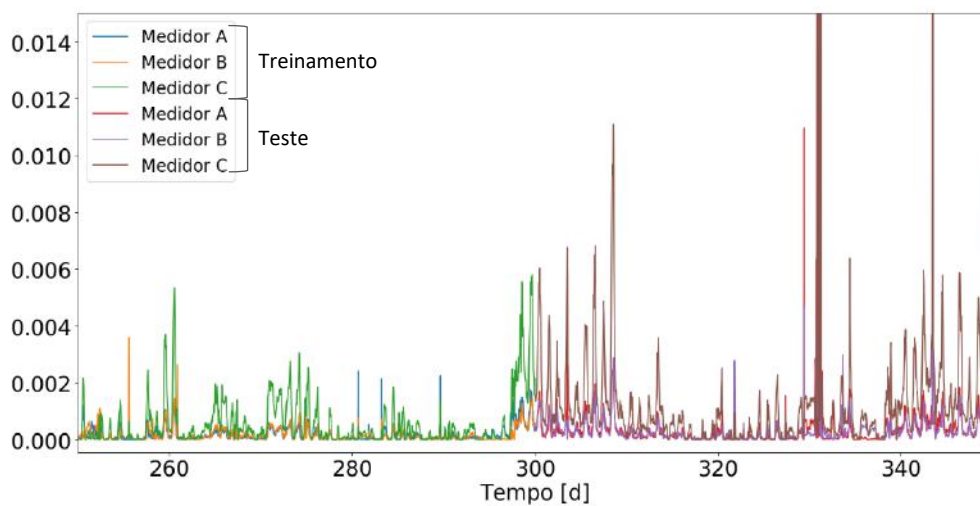


Figura IV-33 Visão das contribuições dos três medidores no erro, com foco no período de transição entre os dados de treinamento e os dados de teste.

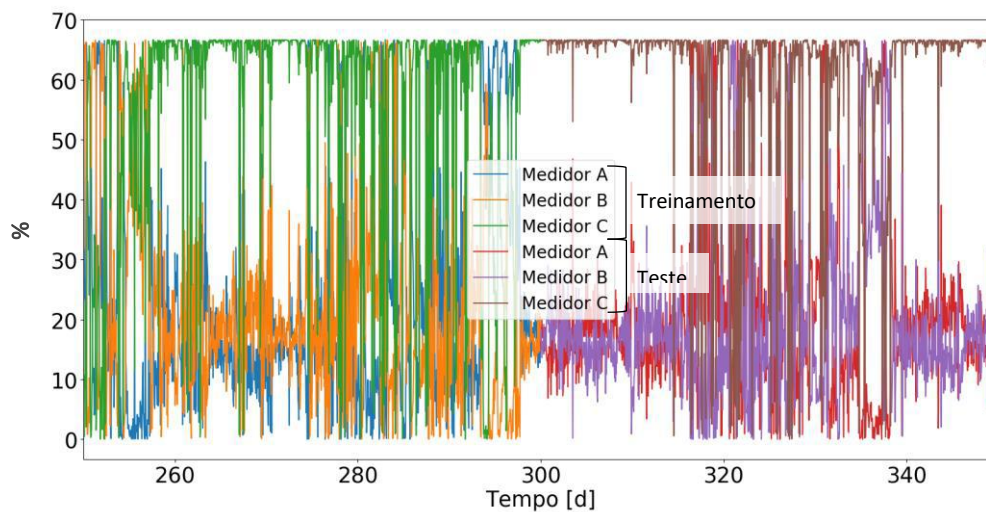


Figura IV-34 Visão do percentual de cada contribuição para o SPE, com foco no período de transição entre os dados de treinamento e os dados de teste.

IV.5 Simulação de erro de medição

Para a comprovação da adequação da técnica de PCA para fins de detecção de erros de medição, foi imputado um erro na base de dados (a partir do dia 317, foram adicionadas 2 t/h aos dados de vazão do medidor A, que corresponde a cerca de 2% da vazão original). Ou seja, os gráficos a seguir foram gerados utilizando dados para treinamento e na fase de teste foram utilizados os dados de processo e a partir de determinado momento, os dados com o erro conhecido.

Percebe-se que o erro de medição proposital do Medidor A foi claramente evidenciado através da terceira variável latente, que anteriormente ao erro, ficava praticamente constante em um patamar e no momento da inserção do erro, mudou para outro patamar (Figuras IV-35, IV-36, IV-37 e IV-38). Mas o SPE não ultrapassou o limite estipulado constantemente durante o período de erro, apesar de ter mudado para um patamar mais elevado (Figuras IV-40 e IV-41). Isso pode mostrar que para fins de monitoramento de erros mais sutis, o limite conforme calculado neste trabalho não consegue atender bem a essa solicitação. Mas, por outro lado, há um evidente aumento no patamar do erro, que poderia ser percebido pelo acompanhamento visual do SPE.

Além disso, foi possível destacar que mesmo com um erro relativo à criação de um *bias* em um dos medidores, a variável latente 2 continuou computando o resíduo relacionado à temperatura ambiente (Figuras IV-37 e IV-38). Comportamento este esperado e desejado para a boa atuação da técnica no problema.

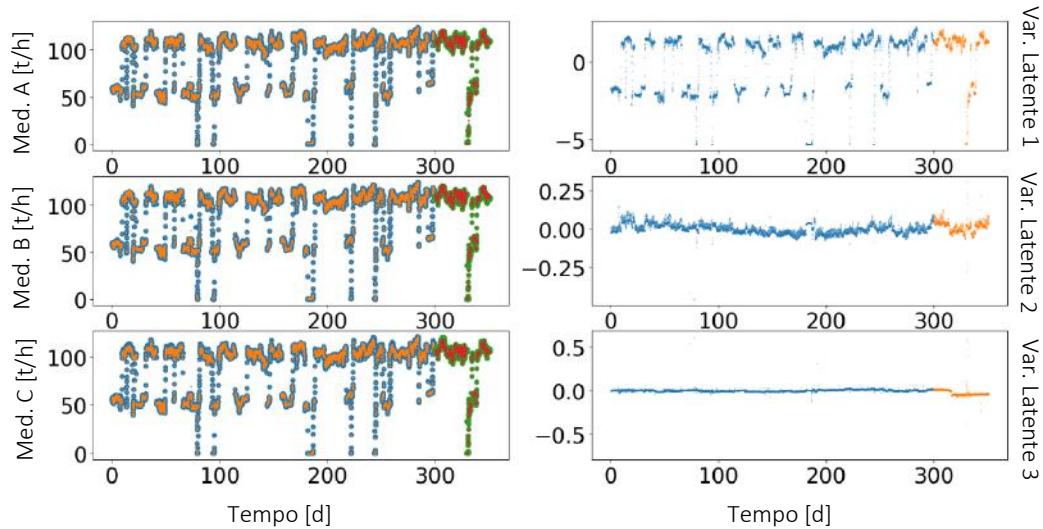


Figura IV-35 Visão inicial ao longo do tempo (dias) das três variáveis observadas/estimadas (lado esquerdo) – em t/h – e das três variáveis latentes (lado direito). O erro foi imputado no Medidor A a partir do dia 317.

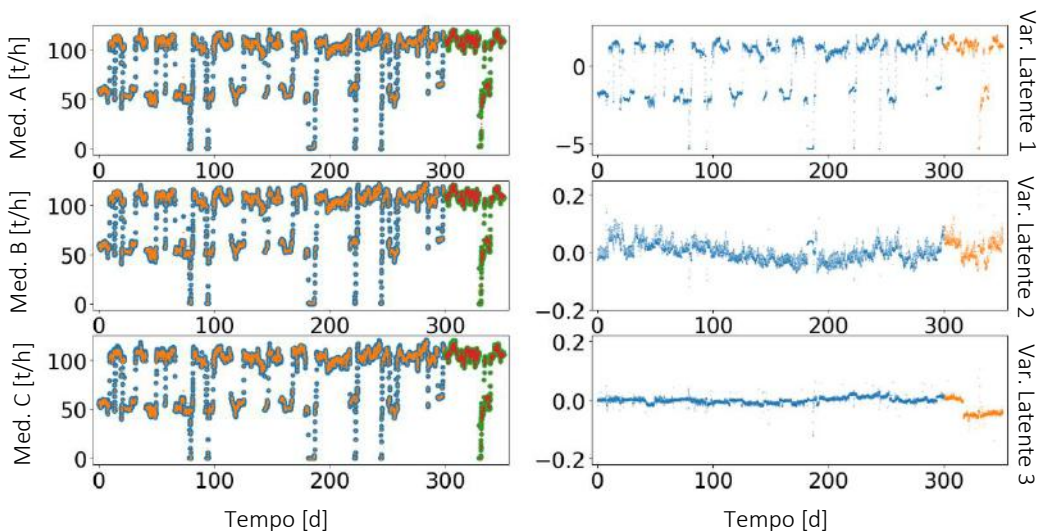


Figura IV-36 Visão ao longo do tempo (dias) das três variáveis observadas/estimadas (lado esquerdo) – em t/h – e das três variáveis latentes (lado direito). O erro foi imputado nos dados para teste do Medidor A a partir do dia 317.

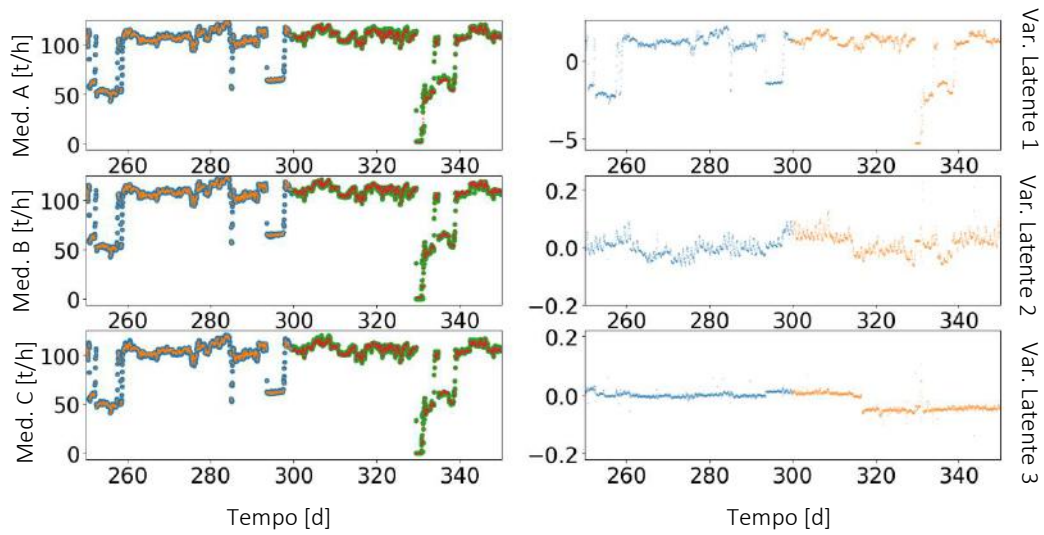


Figura IV-37 Visão ao longo do tempo (dias) das três variáveis observadas/estimadas (lado esquerdo) – em t/h – e das três variáveis latentes (lado direito), com foco no período de transição entre os dados de treinamento e os dados de teste (a partir do dia 300), imputando um erro no Medidor A a partir do dia 317.

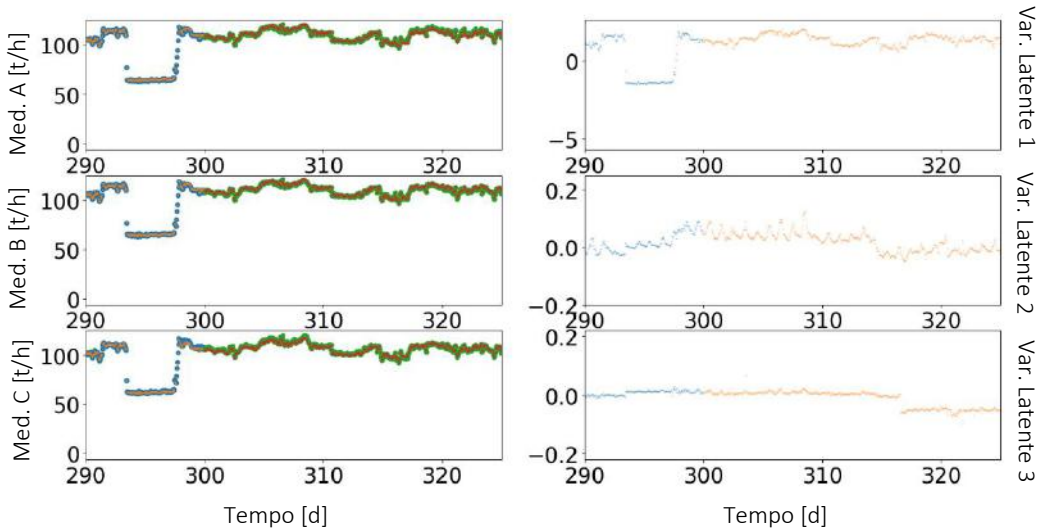


Figura IV-38 Visão ao longo do tempo (dias) das três variáveis observadas/estimadas (lado esquerdo) – em t/h – e das três variáveis latentes (lado direito), com foco no período de transição entre os dados de treinamento e os dados de teste (a partir do dia 300), imputando um erro no Medidor A a partir do dia 317.

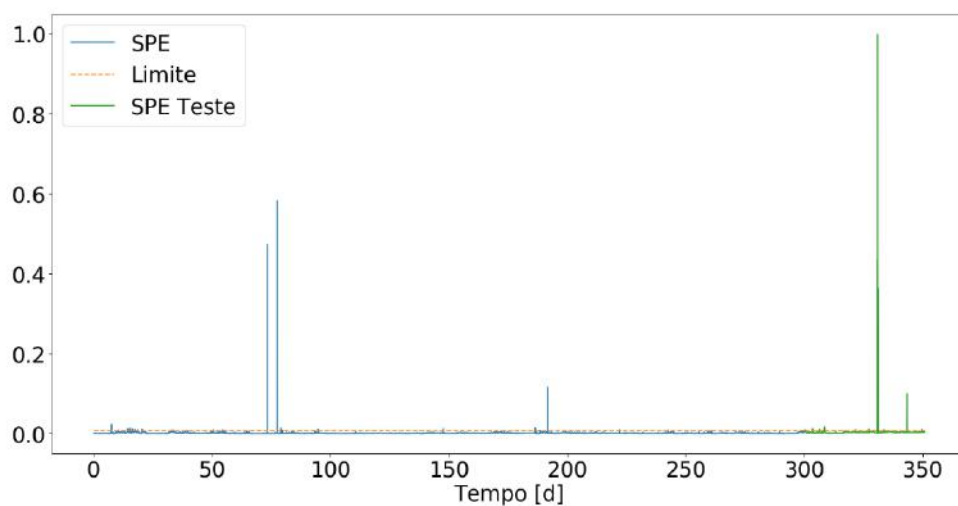


Figura IV-39 Visão inicial ao longo do tempo (dias) das contribuições dos três medidores no erro, utilizando-se dados para treinamento e dados para teste (a partir do dia 300), com erro conhecido no Medidor A a partir do dia 317.

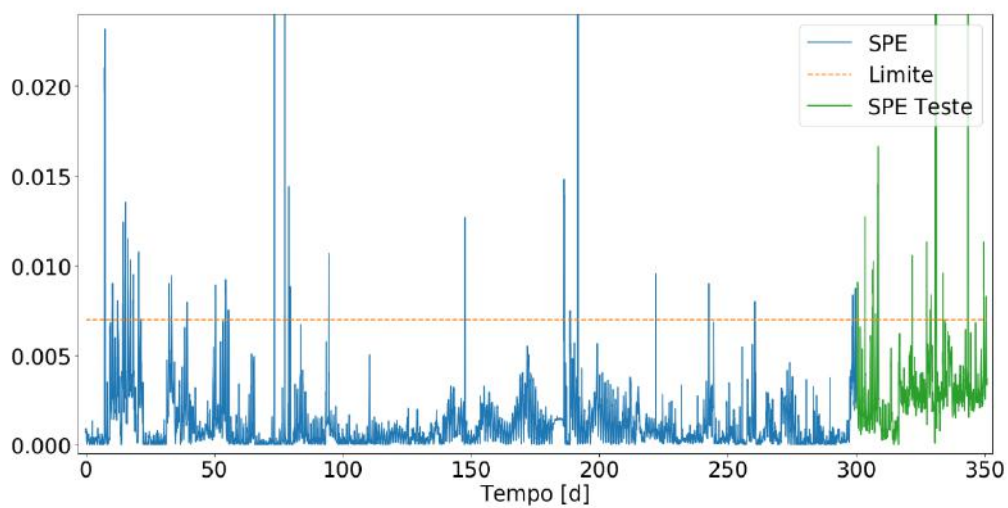


Figura IV-40 Visão ao longo do tempo (dias) das contribuições dos três medidores no erro, utilizando-se dados para treinamento e dados para teste (a partir do dia 300), com erro conhecido no Medidor A a partir do dia 317.

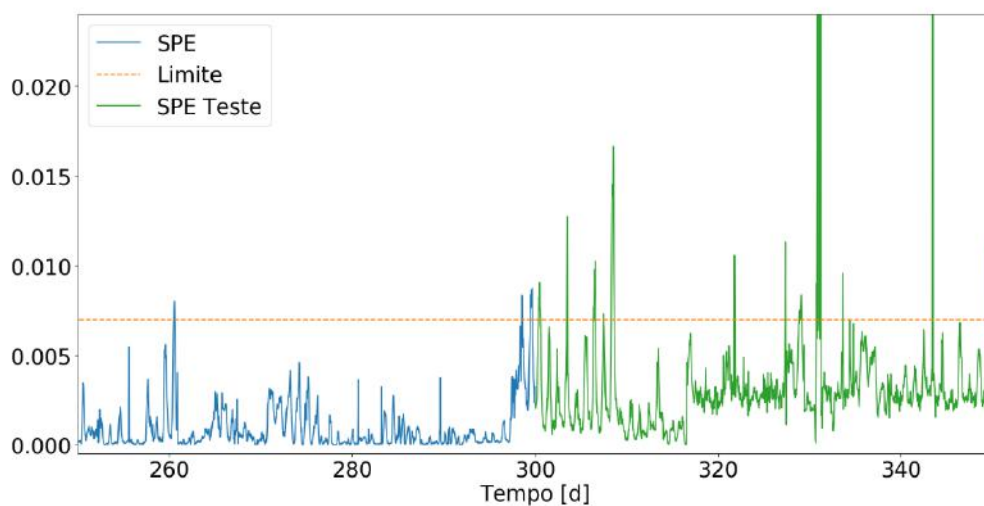


Figura IV-41 Visão das contribuições dos três medidores no erro, com foco na transição entre os dados para treinamento e os dados para teste (a partir do dia 300), com erro conhecido no Medidor A a partir do dia 317.

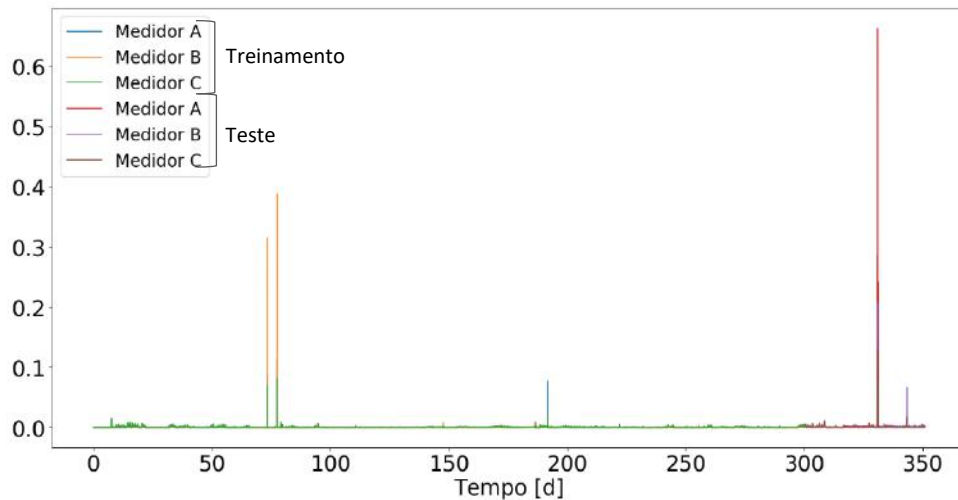


Figura IV-42 Visão inicial ao longo do tempo (dias) das contribuições dos três medidores no erro, utilizando-se dados para treinamento e dados para teste (a partir do dia 300), com erro conhecido no Medidor A a partir do dia 317.

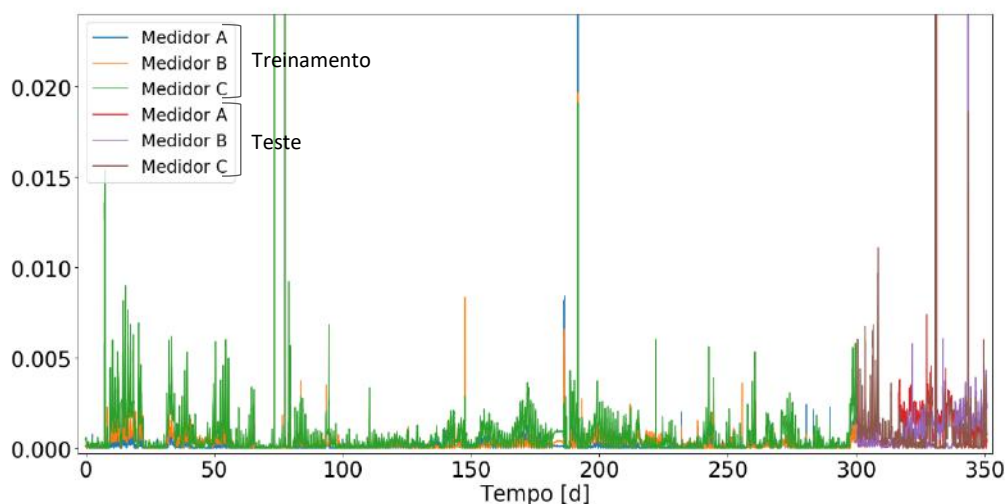


Figura IV-43 Visão ao longo do tempo (dias) das contribuições dos três medidores no erro, utilizando-se dados para treinamento e dados para teste (a partir do dia 300), com erro conhecido no Medidor A a partir do dia 317.

Em relação à análise das contribuições, o *output* parece ter sido satisfatório inicialmente, tendo em vista que a maior contribuição passou a ser do Medidor A (onde foi imputado o erro) e antes era praticamente o tempo todo do medidor C (Figuras IV-44 e IV-46). Mas, a partir de determinado momento, a análise das contribuições passou a indicar que o medidor B estava apresentando erro de medição; isso pode ter acontecido em função de o erro imputado no medidor A ter ocasionalmente sido na mesma magnitude e direção que um erro já presente nos dados coletados do Medidor C (tendo em vista as análises dos dados sem o *input* de erros mapeados, analisado na seção anterior), o que acabou resultando na indicação de um erro no Medidor B, quanto que na verdade o erro estava contido nos medidores A e C.

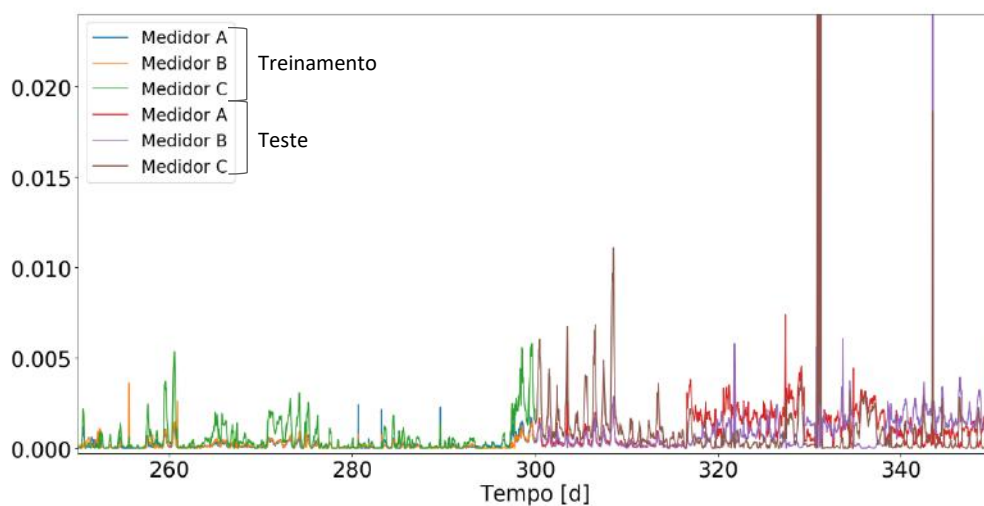


Figura IV-44 Visão das contribuições dos três medidores no erro, com foco na transição entre os dados para treinamento e dados para teste (a partir do dia 300), com erro conhecido no Medidor A a partir do dia 317.

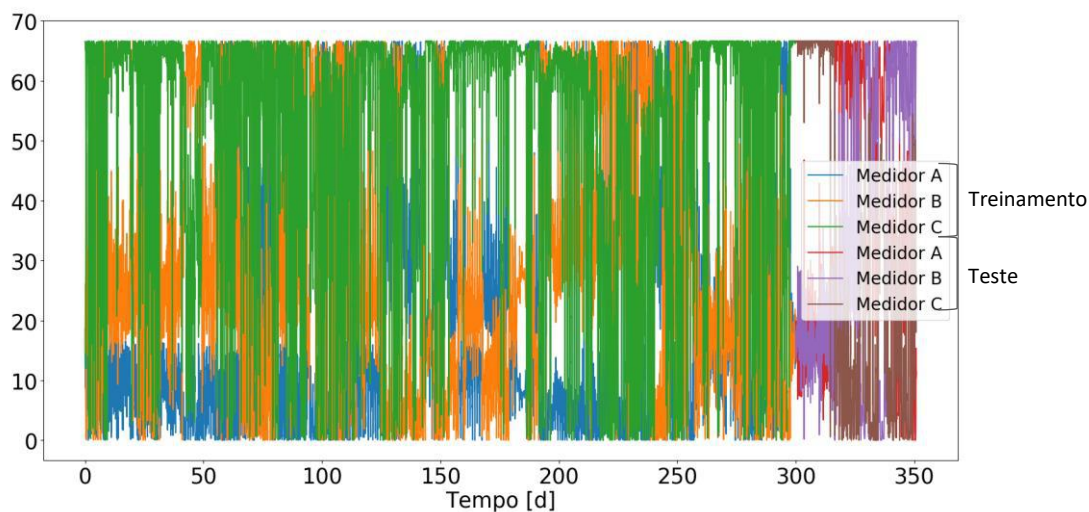


Figura IV-45 Visão ao longo do tempo (dias) do percentual de cada contribuição para o SPE, utilizando-se dados para treinamento e dados para teste (a partir do dia 300), com erro no Medidor A imputado a partir do dia 317.

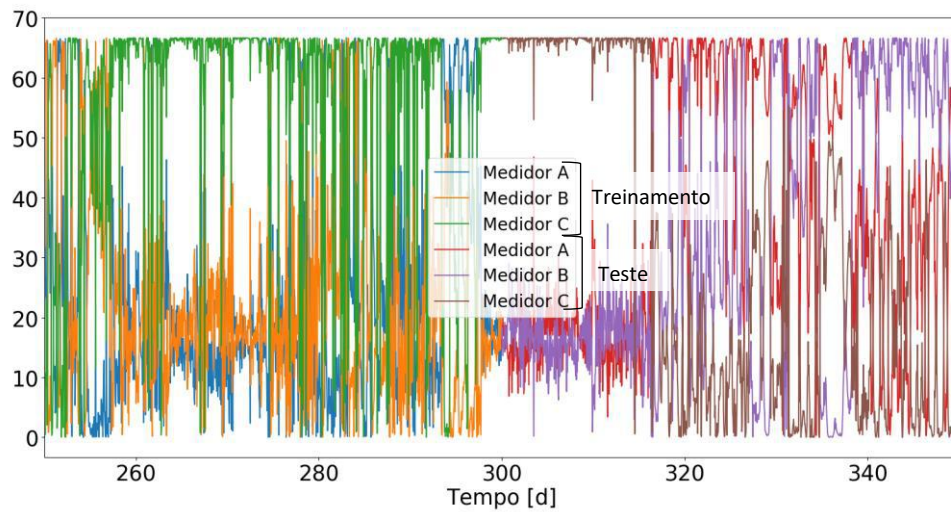


Figura IV-46 Visão do percentual de cada contribuição para o SPE, com foco na transição entre os dados para treinamento e dados para teste (a partir do dia 300), com erro conhecido no Medidor A a partir do dia 317.

Vale destacar que, na prática, não é gerado um gráfico de acompanhamento do percentual da contribuição do medidor no SPE. Normalmente os programas indicam que há erro através das regras de ultrapassagem do limite do SPE e a qualificação de onde vem o erro ocorre com a criação de um gráfico em barras, que determina a contribuição, somente no momento de erro, das variáveis analisadas.

CAPÍTULO V - CONCLUSÃO

V.1 Conclusões

Este trabalho utilizou 2 técnicas estatísticas para a análise de erros de medição em 3 medidores de vazão redundantes, a fim de possibilitar a detecção desses eventuais erros.

A técnica de Espectro de Variância se mostrou importante para a qualificação dos medidores, tendo em vista que com ela foi possível determinar quanto seria o impacto de erros de alta frequência nas medições indicadas. Mas a técnica não se mostrou apropriada para indicar a detecção de novos erros que surgem ao longo do tempo.

Utilizando o espectro de variância, os erros de alta frequência indicados para os Medidores B e C foram estatisticamente iguais, não possibilitando realizar uma diferenciação entre eles. Já o desvio padrão para janela de tamanho 2 do Medidor A ficou um pouco acima dos outros dois instrumentos quando utilizada a base com dados minuto-a-minuto. A partir deste valor, foi possível calcular uma estimativa do impacto econômico do erro. Esse seria o impacto de somente uma corrente de produto; entendendo que o processo possui diversas correntes de produtos e matérias-primas, o impacto cumulativo desses pequenos erros de medição de todas as correntes pode ter uma magnitude bastante elevada, tanto desfavorecendo quanto favorecendo economicamente a empresa e os ecoindicadores, quando na verdade o que se busca são os valores mais próximos possíveis das quantidades reais, para que se tenha pleno conhecimento e controle do processo e visão do negócio.

Pela técnica de PCA, destacou-se o impacto da temperatura ambiente nos valores de vazão medidos pelos instrumentos. Fica então como sugestão a implementação de alguma ação na área física dos medidores, para minimizar o efeito da temperatura ambiente.

A técnica de PCA provou-se satisfatória para detectar inclusive erros de medição mais suaves, aqui simulado um erro de medição de cerca de 2% na medição (2 t/h em uma vazão que originalmente era observada como sendo cerca de 100 t/h). Isso porque foi possível observar nos gráficos, tanto da variável latente 3 quanto no gráfico de SPE, que os respectivos patamares mudaram justamente no instante da fixação do erro. Além disso, a contribuição durante esse período de erro simulado, passou a ser maior por parte do Medidor A (enquanto antes quase se totalizava no Medidor C), até o momento em que passa a indicar que o erro

está contido no medidor B. Com o conhecimento da base de dados, é possível inferir que, na verdade, o erro imputado propositalmente no Medidor A acabou sendo de acordo com um erro presente no medidor C nos últimos dias analisados, o que fez com que o diagnóstico tenha apontado erro no Medidor B, quando na verdade o erro estava presente nos Medidores A e C ao mesmo tempo. Percebe-se que o medidor C normalmente apresenta o maior impacto nos erros acima do limite, por isso deveria ser realizado um estudo um pouco mais aprofundado sobre o Medidor C, de forma a entender seu comportamento, visando criar ações para seu maior controle, objetivando um diagnóstico de falhas mais eficiente para o caso apresentado.

Apesar dessa inconveniência, foi possível comprovar que a técnica permite um diagnóstico de falhas, indicando qual dos medidores de mesma corrente precisam ser verificados/corrigidos. O maior ponto de atenção deve ser para o limite do SPE, que não se mostrou satisfatório para a detecção dos erros mais sutis. Ou seja, antes de se realizar qualquer tipo de carta de controle que sugestione a verificação de algum dos medidores, é necessário repensar o limite do SPE que deverá ser utilizado. Fica também como ponto de atenção a necessidade de uma análise crítica do diagnóstico de falhas, tendo em vista a possibilidade de 2 instrumento falharem na mesma magnitude e direção, o que acaba fazendo com que o diagnóstico indique erro no único instrumento correto.

V.2 Sugestões para Trabalhos Futuros

A princípio, sugere-se a criação de uma lógica para a detecção de falhas, com exploração de outros cálculos de limite para o SPE e determinação da frequência de pontos fora do limite que seriam considerados aceitáveis. E, então, a criação de código para acompanhamento prático na indústria.

Além disso, a aplicação das técnicas em outros problemas práticos relacionados à medição poderia validar ainda mais o procedimento como forma de monitoramento da condição dos medidores na indústria, além de abordar outros problemas práticos.

Sugere-se também trabalhar os erros de medição com outras variáveis, como temperatura e composição de corrente, por exemplo. Isso teria potencial para resolver outros problemas práticos, como, por exemplo, o valor verdadeiro da vazão de uma corrente de

matéria-prima, tendo em vista que a composição muda com frequência e considerando que alguns dos medidores sofrem com a o efeito da mudança do valor da composição.

REFERÊNCIAS

Assis, A.A.; Munaro, C.J.; Rauber, T.W.; Ciarelli, P.M.; Santi, G.B. **Parametrização de diagnóstico de falhas de processos em múltiplas regiões de operação utilizando análise externa**. XIII Simpósio Brasileiro de Automação Inteligente, p 1623-1628. 2017.

Barbosa, A. G. **Desenvolvimento de um Software para Detecção de Erros Grosseiros e Reconciliação de Dados Estática e Dinâmica de Processos Químicos e Petroquímicos**. Tese de Doutorado, Programa de Pós-Graduação em Engenharia Química, Faculdade de Engenharia Química, UNICAMP, 2008.

Bastidas, M.E.H. **Detecção e diagnóstico de falhas baseado em modelos empíricos no subespaço das variáveis de processo (empvsub)**. Dissertação de Mestrado, Programa de Pós-Graduação em Engenharia Química, Escola de Engenharia, UFRGS, 2018.

Feital, T.S. **Monitoramento da condição de processos químicos industriais**. Tese de Doutorado, Programa de Pós-Graduação em Engenharia Química, COPPE, UFRJ, 2011.

Feital, T.; Pinto, J.C.; Prata, D.M. **Estimação das variâncias dos erros de medição em processos químicos industriais**. XV Encontro de Modelagem Computacional. 2012.

Feital, T.; Pinto, J.C. **Use of variance spectra for in-line validation of process measurements in continuous processes**. The Canadian Journal of Chemical Engineering, vol 93, p 1426–1437. 2015.

Ge, Z.; Song, Z. **Distributed PCA Model for Plant-Wide Process Monitoring**. Industrial & Engineering Chemistry Research, vol 52, p 1947-1957. 2013.

Harkat, M.F.; Mourot, G.; Ragot, J. **An improved PCA scheme for sensor FDI: Application to an air quality monitoring network**. Journal of Process Control, vol 16, p 625-634. 2006.

Krishnannair, S.; Aldrich, C. **Fault detection in the Tennessee Eastman benchmark process with nonlinear singular spectrum analysis**. International Federation of Automatic Control, vol 50-1, p 8005–8010. 2017.

Lerner, D. **Uso de modelos de redes neurais artificiais para detecção de falhas no processo Tennessee Eastman.** Dissertação de Mestrado, Programa de Pós-Graduação em Engenharia de Materiais e Processos Químicos e Metalúrgicos, Departamento de Engenharia Química e de Materiais, PUC-Rio, 2018.

Loca, A.L.S.; Rauber, T.W. **Uso de uma rede neural convolucional unidimensional para detecção de falhas em processos industriais.** Programa de Pós-Graduação em Informática, UFES, 2019.

Marotta, G.S.; Silva, A.S.; Assis, L.C.; Andrade, R.J.O.; Gonçalves, R.P. **Identificação de erros grosseiros utilizando teste qui-quadrado e teste tau.** V Colóquio Brasileiro de Ciências Geodésicas, p 393-399. 2007.

Mello, J.M.P. **Consolidação do setor petroquímico brasileiro avaliando a Braskem.** Monografia de final de curso, Departamento de Economia, PUC-RJ, 2017.

Mustafa, G.; Souza, F. **Avaliação preliminar do uso de coluna desmetanizadora complementar para o aumento da capacidade de planta petroquímica.** Seminário Estudantil de Produção Acadêmica, UNIFACS, 2017.

O Setor Petroquímico. Braskem. Disponível em: <<https://www.braskem.com.br/RI/o-setor-petroquimico>>. Acesso em: 21 jan. 2020.

Perfil 2017. Abiplast, 2018. Disponível em: <<http://file.abiplast.org.br/file/download/2018/Perfil-2017.pdf>>. Acesso em: 21 jan. 2020.

Portnoy, I.; Melendez, K.; Pinzon, H.; Sanjuan, M. **An improved weighted recursive PCA algorithm for adaptive fault detection.** Control Engineering Practice, vol 50, p 69-83. 2016.

Qin, S.J. **Statistical process monitoring: basics and beyond.** Journal of Chemometrics, vol 17, p 480-502. 2003.

Qin, S.J.; Yue, H.; Dunia, R. **Self-validating inferential sensors with application to air emission monitoring.** Industrial & Engineering Chemistry Research, vol 36, p 1675-85. 1997.

Shang, L.; Liu, J.; Zhang, Y. **Recursive fault detection and identification for time-varying processes.** Industrial & Engineering Chemistry Research, vol 55, p 12149–12160. 2016.

Tong, C.; Lan, T.; Yu, H.; Shi, X. **Decentralized modified autoregressive models for fault detection in dynamic processes.** Industrial & Engineering Chemistry Research, vol 57, p 15794-15802. 2018.

Van Hombeeck, M. **Comparação entre as rotas via nafta e via etanol para a produção de eteno – análise de custos e emissões de CO₂.** Dissertação de Mestrado, Programa de Pós-Graduação em Planejamento Energético, COPPE, UFRJ, 2019.

Venkatasubramanian, V.; Rengaswamy, R.; Yin, K.; Kavuri, S.N. **A review of process fault detection and diagnosis Part I: Quantitative model-based methods.** Computers and Chemical Engineering, vol 27, p 293-311. 2003.

Yin, S.; Ding, S.X.; Haghani, A.; Hao, H.; Zhang, P. **A comparison study of basic data-driven fault diagnosis and process monitoring methods on the benchmark Tennessee Eastman process.** Journal of Process Control, vol 22, p 1567-1581. 2012.

Yin, J.; Yan, X. **Mutual information–dynamic stacked sparse autoencoders for fault detection.** Industrial & Engineering Chemistry Research, vol 58, p 21614–21624. 2019.

Yoon, S.; MacGregor, J.F. **Fault diagnosis with multivariate statistical models part I: using steady state fault signatures.** Journal of Process Control, vol 11, p 387-400. 2001.

Zhang, Y.; An, J.; Ma, C. **Process monitoring of multimode processes using Kernel Independent Component Analysis.** 8th IFAC Symposium on Advanced Control of Chemical Processes, p 697-701. 2012.

Zhan, C.; Li, S.; Yang, Y. **Enhanced fault detection based on ensemble global–local preserving projections with quantitative global–local structure analysis.** Industrial & Engineering Chemistry Research, vol 56, p 10743–10755. 2017.